

-- 52 --

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: MARIA DIAZ-TORRES
NIGEL DUNN-COLEMAN
MATTHEW CHASE
- (ii) TITLE OF INVENTION: METHOD FOR THE
RECOMBINANT PRODUCTION OF 1,3 PROPANEDIOL
- (iii) NUMBER OF SEQUENCES: 49
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: GENENCOR INTERNATIONAL, INC.
 - (B) STREET: 4 CAMBRIDGE PLACE
1870 SOUTH WINTON ROAD
 - (C) CITY: ROCHESTER
 - (D) STATE: NEW YORK
 - (E) COUNTRY: U.S.A.
 - (F) POSTAL CODE (ZIP): 14618
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: 3.50 INCH DISKETTE
 - (B) COMPUTER: IBM PC COMPATIBLE
 - (C) OPERATING SYSTEM: MICROSOFT WINDOWS 3.1
 - (D) SOFTWARE: MICROSOFT WORD 2.0C
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE: 11/13/97
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: 60/030,601
 - (B) FILING DATE: 11/13/96
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: GLAISTER, DEBRA
 - (B) REGISTRATION NO.: 33,888
 - (C) REFERENCE/DOCKET NUMBER: GC 369-2
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 650-864-7620
 - (B) TELEFAX: 650-845-6504

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1668 base pairs

-- 53 --

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: DHAB1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGAAAAGAT CAAAACGATT TGCAGTACTG GCCCAGCGCC CCGTCAATCA GGACGGGCTG	60
ATTGGCGAGT GGCCTGAAGA GGGGCTGATC GCCATGGACA GCCCCTTTGA CCCGGTCTCT	120
TCAGTAAAAG TGGACAACGG TCTGATCGTC GAACTGGACG GCAAACGCCG GGACCAGTTT	180
GACATGATCG ACCGATTTAT CGCCGATTAC GCGATCAACG TTGAGCGCAC AGAGCAGGCA	240
ATGCGCCTGG AGGCGGTGGA AATAGCCCGT ATGCTGGTGG ATATTCACGT CAGCCGGGAG	300
GAGATCATTG CCATCACTAC CGCCATCACG CCGGCCAAAG CGGTCGAGGT GATGGCGCAG	360
ATGAACGTGG TGGAGATGAT GATGGCGCTG CAGAAGATGC GTGCCCCCGG GACCCCCTCC	420
AACCAAGTGC ACGTACACAA TCTCAAAGAT AATCCGGTGC AGATTGCCGC TGACGCCGCC	480
GAGGCCGGGA TCCGCGGCTT CTCAGAACAG GAGACCACGG TCGGTATCGC GCGCTACGCG	540
CCGTTTAACG CCCTGGCGCT GTTGGTCGGT TCGCAGTGC GCGCCCCCGG CGTGTGACG	600
CAGTGCTCGG TGGAAGAGGC CACCGAGCTG GAGCTGGGCA TGCGTGGCTT AACCAGCTAC	660
GCCGAGACGG TGTCGGTCTA CGGCACCGAA GCGGTATTTA CCGACGGCGA TGATACGCCG	720
TGGTCAAAGG CGTTCCTCGC CTCGGCCTAC GCCTCCCGCG GGTGAAAAT GCGCTACACC	780
TCCGGCACCG GATCCGAAGC GCTGATGGGC TATTCGGAGA GCAAGTCGAT GCTCTACCTC	840
GAATCGCGCT GCATCTTCAT TACTAAAGGC GCCGGGGTTC AGGGACTGCA AAACGGCGCG	900
GTGAGCTGTA TCGGCATGAC CGGCGCTGTG CCGTCGGGCA TTCGGGCGGT GCTGGCGGAA	960
AACCTGATCG CCTCTATGCT CGACCTCGAA GTGGCGTCCG CCAACGACCA GACTTTCTCC	1020
CACTCGGATA TTCGCCGCAC CGCGCGCACC CTGATGCAGA TGCTGCCGGG CACCGACTTT	1080
ATTTTCTCCG GCTACAGCGC GGTGCCGAAC TACGACAACA TGTTCCGCCG CTCGAACTTC	1140
GATGCGGAAG ATTTTGATGA TTACAACATC CTGCAGCGTG ACCTGATGGT TGACGGCGGC	1200
CTGCGTCCGG TGACCGAGGC GGAAACCATT GCCATTCGCC AGAAAGCGGC GCGGGCGATC	1260

-- 54 --

CAGGCGGTTT TCCGCGAGCT GGGGCTGCCG CCAATCGCCG ACGAGGAGGT GGAGGCCGCC 1320
 ACCTACGCGC ACGGCAGCAA CGAGATGCCG CCGCGTAACG TGGTGGAGGA TCTGAGTGCG 1380
 GTGGAAGAGA TGATGAAGCG CAACATCACC GGCCTCGATA TTGTCGGCGC GCTGAGCCGC 1440
 AGCGGCTTTG AGGATATCGC CAGCAATATT CTCAATATGC TGCGCCAGCG GGTACCCGGC 1500
 GATTACCTGC AGACCTCGGC CATTCTCGAT CGGCAGTTCG AGGTGGTGAG TGCGGTCAAC 1560
 GACATCAATG ACTATCAGGG GCCGGGCACC GGCTATCGCA TCTCTGCCGA ACGCTGGGCG 1620
 GAGATCAAAA ATATTCCGGG CGTGGTTCAG CCCGACACCA TTGAATAA 1668

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 585 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: DHAB2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GTGCAACAGA CAACCCAAAT TCAGCCCTCT TTTACCCTGA AAACCCGCGA GGGCGGGGTA 60
 GCTTCTGCCG ATGAACGCGC CGATGAAGTG GTGATCGGCG TCGGCCCTGC CTTGATAAA 120
 CACCAGCATC AACTCTGAT CGATATGCCC CATGGCGCGA TCCTCAAAGA GCTGATTGCC 180
 GGGGTGGAAG AAGAGGGGCT TCACGCCCGG GTGGTGCGCA TTCTGCGCAC GTCCGACGTC 240
 TCCTTTATGG CCTGGGATGC GGCCAACCTG AGCGGCTCGG GGATCGGCAT CGGTATCCAG 300
 TCGAAGGGGA CCACGGTCAT CCATCAGCGC GATCTGCTGC CGCTCAGCAA CCTGGAGCTG 360
 TTCTCCCAGG CGCCGCTGCT GACGCTGGAG ACCTACCGGC AGATTGGCAA AAACGCTGCG 420
 CGCTATGCGC GCAAAGAGTC ACCTTCGCCG GTGCCGGTGG TGAACGATCA GATGGTGCGG 480
 CCGAAATTTA TGGCCAAAGC CGCGCTATTT CATATCAAAG AGACCAAACA TGTGGTGCGG 540
 GACGCCGAGC CCGTCACCCT GCACATCGAC TTAGTAAGGG AGTGA 585

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 426 base pairs
- (B) TYPE: nucleic acid

-- 55 --

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: DHAB3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATGAGCGAGA AAACCATGCG CGTGCAGGAT TATCCGTTAG CCACCCGCTG CCCGGAGCAT    60
ATCCTGACGC CTACCGGCAA ACCATTGACC GATATTACCC TCGAGAAGGT GCTCTCTGGC    120
GAGGTGGGCC CGCAGGATGT GCGGATCTCC CGCCAGACCC TTGAGTACCA GGC GCAGATT    180
GCCGAGCAGA TGCAGCGCCA TGC GGTGGCG CGCAATTTCC GCCGCGCGGC GGAGCTTATC    240
GCCATTCCTG ACGAGCGCAT TCTGGCTATC TATAACGCGC TGCGCCCGTT CCGCTCCTCG    300
CAGGCGGAGC TGCTGGCGAT CGCCGACGAG CTGGAGCACA CCTGGCATGC GACAGTGAAT    360
GCCGCCTTTG TCCGGGAGTC GGC GGAAGTG TATCAGCAGC GGCATAAGCT GCGTAAAGGA    420
AGCTAA                                           426
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1164 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: DHAT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
ATGAGCTATC GTATGTTTGA TTATCTGGTG CCAAACGTTA ACTTTTTTTGG CCCCAACGCC    60
ATTTCCGTAG TCGGCGAACG CTGCCAGCTG CTGGGGGGGA AAAAAGCCCT GCTGGTCACC    120
GACAAAGGCC TCGGGGCAAT TAAAGATGGC GCGGTGGACA AAACCCTGCA TTATCTGCGG    180
GAGGCCGGGA TCGAGGTGGC GATCTTTGAC GCGGTCGAGC CGAACCCGAA AGACACCAAC    240
GTGCGCGACG GCCTCGCCGT GTTTCGCCGC GAACAGTGCG ACATCATCGT CACCGTGGGC    300
GGCGGCAGCC CGCACGATTG CGGCAAAGGC ATCGGCATCG CCGCCACCCA TGAGGGCGAT    360
CTGTACCAGT ATGCCGGAAT CGAGACCCTG ACCAACCCGC TGCCGCCTAT CGTCGCGGTC    420
AATACCACCG CCGGCACCGC CAGCGAGGTC ACCCGCCACT GCGTCCTGAC CAACACCGAA    480
```

-- 56 --

ACCAAAGTGA AGTTTGTGAT CGTCAGCTGG CGCAAACCTGC CGTCGGTCTC TATCAACGAT 540
 CCACTGCTGA TGATCGGTAA ACCGGCCGCC CTGACCGCGG CGACCGGGAT GGATGCCCTG 600
 ACCCAGCCCG TAGAGGCCTA TATCTCCAAA GACGCTAACC CGGTGACGGA CGCCGCCGCC 660
 ATGCAGGCGA TCCGCCTCAT CGCCCGCAAC CTGCGCCAGG CCGTGCCCTT CGGCAGCAAT 720
 CTGCAGGCGC GGGAAAACAT GGCCTATGCT TCTCTGCTGG CCGGGATGGC TTTCAATAAC 780
 GCCAACCTCG GCTACGTGCA CGCCATGGCG CACCAGCTGG GCGGCCTGTA CGACATGCCG 840
 CACGGCGTGG CCAACGCTGT CCTGCTGCCG CATGTGGCGC GCTACAACCT GATCGCCAAC 900
 CCGGAGAAAT TCGCCGATAT CGCTGAACTG ATGGGCGAAA ATATCACCGG ACTGTCCACT 960
 CTCGACGCGG CGGAAAAAGC CATCGCCGCT ATCACGCGTC TGTCGATGGA TATCGGTATT 1020
 CCGCAGCATC TGCGCGATCT GGGGGTAAAA GAGGCCGACT TCCCCTACAT GCGGGAGATG 1080
 GCTCTAAAAG ACGGCAATGC GTTCTCGAAC CCGCGTAAAG GCAACGAGCA GGAGATTGCC 1140
 GCGATTTTCC GCCAGGCATT CTGA 1164

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1380 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPD1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CTTTAATTTT CTTTATCTT ACTCTCTAC ATAAGACATC AAGAAACAAT TGTATATTGT 60
 ACACCCCCC CCTCCACAAA CACAAATATT GATAATATAA AGATGTCTGC TGCTGCTGAT 120
 AGATTAAACT TAACTTCCGG CCACTTGAAT GCTGGTAGAA AGAGAAGTTC CTCTTCTGTT 180
 TCTTTGAAGG CTGCCGAAAA GCCTTTCAAG GTTACTGTGA TTGGATCTGG TAACTGGGGT 240
 ACTACTATTG CCAAGGTGGT TGCCGAAAAT TGTAAGGGAT ACCCAGAAGT TTTCGCTCCA 300
 ATAGTACAAA TGTGGGTGTT CGAAGAAGAG ATCAATGGTG AAAAATTGAC TGAAATCATA 360
 AATACTAGAC ATCAAAACGT GAAATACTTG CCTGGCATCA CTCTACCCGA CAATTTGGTT 420
 GCTAATCCAG ACTTGATTGA TTCAGTCAAG GATGTCGACA TCATCGTTTT CAACATTCCA 480

-- 57 --

CATCAATTTT TGCCCCGTAT CTGTAGCCAA TTGAAAGGTC ATGTTGATTC ACACGTCAGA 540
GCTATCTCCT GTCTAAAGGG TTTTGAAGTT GGTGCTAAAG GTGTCCAATT GCTATCCTCT 600
TACATCACTG AGGAACTAGG TATTCAATGT GGTGCTCTAT CTGGTGCTAA CATTGCCACC 660
GAAGTCGCTC AAGAACACTG GTCTGAAACA ACAGTTGCTT ACCACATTCC AAAGGATTTT 720
AGAGGCGAGG GCAAGGACGT CGACCATAAG GTTCTAAAGG CCTTGTTCCA CAGACCTTAC 780
TTCCACGTTA GTGTCATCGA AGATGTTGCT GGTATCTCCA TCTGTGGTGC TTTGAAGAAC 840
GTTGTTGCCT TAGGTTGTGG TTTCGTCGAA GGTCTAGGCT GGGGTAACAA CGCTTCTGCT 900
GCCATCCAAA GAGTCGGTTT GGGTGAGATC ATCAGATTCG GTCAAATGTT TTTCCCAGAA 960
TCTAGAGAAG AAACATACTA CCAAGAGTCT GCTGGTGTTG CTGATTTGAT CACCACCTGC 1020
GCTGGTGGA GAAACGTCAA GGTGCTAGG CTAATGGCTA CTTCTGGTAA GGACGCCTGG 1080
GAATGTGAAA AGGAGTTGTT GAATGGCCAA TCCGCTCAAG GTTTAATTAC CTGCAAAGAA 1140
GTTACGAAT GGTGGAAC ATGTGGCTCT GTCGAAGACT TCCCATTATT TGAAGCCGTA 1200
TACCAAATCG TTTACAACAA CTACCCAATG AAGAACCTGC CGGACATGAT TGAAGAATTA 1260
GATCTACATG AAGATTAGAT TTATTGGAGA AAGATAACAT ATCATACTTC CCCCCTTTT 1320
TTCGAGGCTC TTCTATATCA TATTCATAAA TTAGCATTAT GTCATTTCTC ATAACACTT 1380

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2946 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: GPD2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GAATTCGAGC CTGAAGTGCT GATTACCTTC AGGTAGACTT CATCTTGACC CATCAACCCC 60
AGCGTCAATC CTGCAAATAC ACCACCCAGC AGCACTAGGA TGATAGAGAT AATATAGTAC 120
GTGGTAACGC TTGCCTCATC ACCTACGCTA TGGCCGGAAT CGGCAACATC CCTAGAATTG 180
AGTACGTGTG ATCCGGATAA CAACGGCAGT GAATATATCT TCGGTATCGT AAAGATGTGA 240
TATAAGATGA TGTATACCCA ATGAGGAGCG CCTGATCGTG ACCTAGACCT TAGTGGCAAA 300
AACGACATAT CTATTATAGT GGGGAGAGTT TCGTGCAAAT AACAGACGCA GCAGCAAGTA 360

-- 58 --

ACTGTGACGA TATCAACTCT TTTTTTATTA TGTAATAAGC AAACAAGCAC GAATGGGGAA 420
 AGCCTATGTG CAATCACCAA GGTCGTCCCT TTTTTCCTAT TTGCTAATTT AGAATTTAAA 480
 GAAACCAAAA GAATGAAGAA AGAAAACAAA TACTAGCCCT AACCTGACT TCGTTTCTAT 540
 GATAATACCC TGCTTTAATG AACGGTATGC CCTAGGGTAT ATCTCACTCT GTACGTTACA 600
 AACTCCGGTT ATTTTATCGG AACATCCGAG CACCCGCGCC TTCCTCAACC CAGGCACCGC 660
 CCCAGGTAAC CGTGCGCGAT GAGCTAATCC TGAGCCATCA CCCACCCAC CCGTTGATGA 720
 CAGCAATTCG GGAGGGCGAA AATAAACTG GAGCAAGGAA TTACCATCAC CGTCACCATC 780
 ACCATCATAT CGCCTTAGCC TCTAGCCATA GCCATCATGC AAGCGTGTAT CTTCTAAGAT 840
 TCAGTCATCA TCATTACCGA GTTTGTTTTT CTTACATGA TGAAGAAGGT TTGAGTATGC 900
 TCGAAACAAT AAGACGACGA TGGCTCTGCC ATTGGTTATA TTACGCTTTT GCGGCGAGGT 960
 GCCGATGGGT TGCTGAGGGG AAGAGTGTTT AGCTTACGGA CCTATTGCCA TTGTTATTCC 1020
 GATTAATCTA TTGTTCAAGCA GCTCTTCTCT ACCCTGTCAT TCTAGTATTT TTTTTTTTTT 1080
 TTTTTGGTTT TACTTTTTTT TCTTCTTGCC TTTTTTCTT GTTACTTTTT TTCTAGTTTT 1140
 TTTTCCTTCC ACTAAGCTTT TTCCTTGATT TATCCTTGGG TTCTTCTTTC TACTCCTTTA 1200
 GATTTTTTTTT TTATATATTA ATTTTAAAGT TTATGTATTT TGGTAGATTC AATTCTCTTT 1260
 CCCTTTCCTT TTCCTTCGCT CCCCTTCCTT ATCAATGCTT GCTGTCAGAA GATTAACAAG 1320
 ATACACATTC CTTAAGCGAA CGCATCCGGT GTTATATACT CGTCGTGCAT ATAAAATTTT 1380
 GCCTTCAAGA TCTACTTTCC TAAGAAGATC ATTATTACAA ACACAACTGC ACTCAAAGAT 1440
 GACTGCTCAT ACTAATATCA AACAGCACAA AACTGTGCAT GAGGACCATC CTATCAGAAG 1500
 ATCGGACTCT GCCGTGTCAA TTGTACATTT GAAACGTGCG CCCTTCAAGG TTACAGTGAT 1560
 TGGTTCTGGT AACTGGGGGA CCACCATCGC CAAAGTCATT GCGGAAAACA CAGAATTGCA 1620
 TTCCCATATC TTCGAGCCAG AGGTGAGAAT GTGGGTTTTT GATGAAAAGA TCGGCGACGA 1680
 AAATCTGACG GATATCATAA ATACAAGACA CCAGAACGTT AAATATCTAC CCAATATTGA 1740
 CCTGCCCCAT AATCTAGTGG CCGATCCTGA TCTTTTACAC TCCATCAAGG GTGCTGACAT 1800
 CCTTGTTTTT AACATCCCTC ATCAATTTTT ACCAAACATA GTCAAACAAT TGCAAGGCCA 1860
 CGTGGCCCCCT CATGTAAGGG CCATCTCGTG TCTAAAAGGG TTCGAGTTGG GCTCCAAGGG 1920
 TGTGCAATTG CTATCCTCCT ATGTTACTGA TGAGTTAGGA ATCCAATGTG GCGCACTATC 1980
 TGGTGCAAAC TTGGCACCGG AAGTGGCCAA GGAGCATTGG TCCGAAACCA CCGTGGCTTA 2040

-- 59 --

CCAACTACCA AAGGATTATC AAGGTGATGG CAAGGATGTA GATCATAAGA TTTTGAAATT 2100
 GCTGTTCCAC AGACCTTACT TCCACGTCAA TGTCATCGAT GATGTTGCTG GTATATCCAT 2160
 TGCCGGTGCC TTGAAGAACG TCGTGGCACT TGCATGTGGT TTCGTAGAAG GTATGGGATG 2220
 GGGTAACAAT GCCTCCGCAG CCATTCAAAG GCTGGGTTTA GGTGAAATTA TCAAGTTCGG 2280
 TAGAATGTTT TTCCCAGAAT CCAAAGTCGA GACCTACTAT CAAGAATCCG CTGGTGTTGC 2340
 AGATCTGATC ACCACCTGCT CAGGCGGTAG AAACGTCAAG GTTGCCACAT ACATGGCCAA 2400
 GACCGGTAAG TCAGCCTTGG AAGCAGAAAA GGAATTGCTT AACGGTCAAT CCGCCCAAGG 2460
 GATAATCACA TGCAGAGAAG TTCACGAGTG GCTACAAACA TGTGAGTTGA CCCAAGAATT 2520
 CCCAATTATT CGAGGCAGTC TACCAGATAG TCTACAACAA CGTCCGCATG GAAGACCTAC 2580
 CGGAGATGAT TGAAGAGCTA GACATCGATG ACGAATAGAC ACTCTCCCCC CCCCTCCCCC 2640
 TCTGATCTTT CCTGTTGCCT CTTTTTCCCC CAACCAATTT ATCATTATAC ACAAGTTCTA 2700
 CAACTACTAC TAGTAACATT ACTACAGTTA TTATAATTTT CTATTCTCTT TTTCTTTAAG 2760
 AATCTATCAT TAACGTTAAT TTCTATATAT ACATAACTAC CATTATACAC GCTATTATCG 2820
 TTTACATATC ACATCACCGT TAATGAAAGA TACGACACCC TGTACACTAA CACAATTAAA 2880
 TAATCGCCAT AACCTTTTCT GTTATCTATA GCCCTTAAAG CTGTTTCTTC GAGCTTTTCA 2940
 CTGCAG 2946

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3178 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: GUT2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CTGCAGAACT TCGTCTGCTC TGTGCCCATC CTCGCGGTTA GAAAGAAGCT GAATTGTTTC 60
 ATGCGCAAGG GCATCAGCGA GTGACCAATA ATCACTGCAC TAATTCCTTT TTAGCAACAC 120
 ATACTTATAT ACAGCACCAG ACCTTATGTC TTTTCTCTGC TCCGATACGT TATCCCACCC 180
 AACTTTTATT TCAGTTTTGG CAGGGGAAAT TTCACAACCC CGCACGCTAA AAATCGTATT 240

-- 60 --

TAAACTTAAA	AGAGAACAGC	CACAAATAGG	GAAC TTTGGT	CTAAACGAAG	GACTCTCCCT	300
CCCTTATCTT	GACCGTGCTA	TTGCCATCAC	TGCTACAAGA	CTAAATACGT	ACTAATATAT	360
GTTTTCGGTA	ACGAGAAGAA	GAGCTGCCGG	TGCAGCTGCT	GCCATGGCCA	CAGCCACGGG	420
GACGCTGTAC	TGGATGACTA	GCCAAGGTGA	TAGGCCGTTA	GTGCACAATG	ACCCGAGCTA	480
CATGGTGCAA	TTCCCCACCG	CCGCTCCACC	GGCAGGTCTC	TAGACGAGAC	CTGCTGGACC	540
GTCTGGACAA	GACGCATCAA	TTCGACGTGT	TGATCATCGG	TGGCGGGGCC	ACGGGGACAG	600
GATGTGCCCT	AGATGCTGCG	ACCAGGGGAC	TCAATGTGGC	CCTTGTTGAA	AAGGGGGATT	660
TTGCCTCGGG	AACGTCGTCC	AAATCTACCA	AGATGATTCA	CGGTGGGGTG	CGGTACTTAG	720
AGAAGGCCTT	CTGGGAGTTC	TCCAAGGCAC	AACTGGATCT	GGTCATCGAG	GCACTCAACG	780
AGCGTAAACA	TCTTATCAAC	ACTGCCCCTC	ACCTGTGCAC	GGTGCTACCA	ATTCTGATCC	840
CCATCTACAG	CACCTGGCAG	GTCCCGTACA	TCTATATGGG	CTGTAAATTC	TACGATTTCT	900
TTGGCGGTTT	CCAAAACCTG	AAAAAATCAT	ACCTACTGTC	CAAATCCGCC	ACCGTGGAGA	960
AGGCTCCCAT	GCTTACCACA	GACAATTTAA	AGGCCTCGCT	TGTGTACCAT	GATGGGTCCT	1020
TTAACGACTC	GCGTTTGAAC	GCCACTTTAG	CCATCACGGG	TGTGGAGAAC	GGCGCTACCG	1080
TCTTGATCTA	TGTCGAGGTA	CAAAAATTGA	TCAAAGACCC	AACTTCTGGT	AAGGTTATCG	1140
GTGCCGAGGC	CCGGGACGTT	GAGACTAATG	AGCTTGTCAG	AATCAACGCT	AAATGTGTGG	1200
TCAATGCCAC	GGGCCCATAC	AGTGACGCCA	TTTTGCAAAT	GGACCGCAAC	CCATCCGGTC	1260
TGCCGGACTC	CCCCTAAAC	GACAAC TCCA	AGATCAAGTC	GACTTTCAAT	CAAATCTCCG	1320
TCATGGACCC	GAAAATGGTC	ATCCCATCTA	TTGGCGTTCA	CATCGTATTG	CCCTCTTTTT	1380
ACTCCCCGAA	GGATATGGGT	TTGTTGGACG	TCAGAACCTC	TGATGGCAGA	GTGATGTTCT	1440
TTTTACCTTG	GCAGGGCAAA	GTCCTTGCCG	GCACCACAGA	CATCCCACTA	AAGCAAGTCC	1500
CAGAAAACCC	TATGCCTACA	GAGGCTGATA	TTCAAGATAT	CTTGAAAGAA	CTACAGCACT	1560
ATATCGAATT	CCCCGTGAAA	AGAGAAGACG	TGCTAAGTGC	ATGGGCTGGT	GTCAGACCTT	1620
TGGTCAGAGA	TCCACGTACA	ATCCCCGCAG	ACGGGAAGAA	GGGCTCTGCC	ACTCAGGGCG	1680
TGGTAAGATC	CCACTTCTTG	TTCACTTCGG	ATAATGGCCT	AATTACTATT	GCAGGTGGTA	1740
AATGGACTAC	TTACAGACAA	ATGGCTGAGG	AAACAGTCGA	CAAAGTTGTC	GAAGTTGGCG	1800
GATTCCACAA	CCTGAAACCT	TGTCACACAA	GAGATATTAA	GCTTGCTGGT	GCAGAAGAAT	1860
GGACGCAAAA	CTATGTGGCT	TTATTGGCTC	AAAAC TACCA	TTTATCATCA	AAAATGTCCA	1920

-- 61 --

ACTACTTGGT TCAAACTAC GGAACCCGTT CCTCTATCAT TTGCGAATTT TTCAAAGAAT 1980
 CCATGGAAAA TAAACTGCCT TTGTCCTTAG CCGACAAGGA AAATAACGTA ATCTACTCTA 2040
 GCGAGGAGAA CAACTTGGTC AATTTTGATA CTTTCAGATA TCCATTACACA ATCGGTGAGT 2100
 TAAAGTATTC CATGCAGTAC GAATATTGTA GAACTCCCTT GGACTTCCTT TTAAGAAGAA 2160
 CAAGATTGCG CTTCTTGGAC GCCAAGGAAG CTTTGAATGC CGTGCATGCC ACCGTCAAAG 2220
 TTATGGGTGA TGAGTTCAAT TGGTCGGAGA AAAAGAGGCA GTGGGAACTT GAAAAAACTG 2280
 TGAAC TTCAT CCAAGGACGT TTCGGTGTCT AAATCGATCA TGATAGTTAA GGGTGACAAA 2340
 GATAACATTC ACAAGAGTAA TAATAATGGT AATGATGATA ATAATAATAA TGATAGTAAT 2400
 AACAATAATA ATAATGGTGG TAATGGCAAT GAAATCGCTA TTATTACCTA TTTTCCTTAA 2460
 TGGAAGAGTT AAAGTAACT AAAAAACTA CAAAATATA TGAAGAAAAA AAAAAAAGA 2520
 GGTAAATAGAC TCTACTACTA CAATTGATCT TCAAATTATG ACCTTCCTAG TGTTTATATT 2580
 CTATTTCCAA TACATAATAT AATCTATATA ATCATTGCTG GTAGACTTCC GTTTTAATAT 2640
 CGTTTTAATT ATCCCCTTTA TCTCTAGTCT AGTTTTATCA TAAAATATAG AAACACTAAA 2700
 TAATATTCTT CAAACGGTCC TGGTGCATAC GCAATACATA TTTATGGTGC AAAAAAAAAA 2760
 ATGGAAAATT TTGCTAGTCA TAAACCCTTT CATAAAACAA TACGTAGACA TCGCTACTTG 2820
 AAATTTTCAA GTTTTTTATCA GATCCATGTT TCCTATCTGC CTTGACAACC TCATCGTCGA 2880
 AATAGTACCA TTTAGAACGC CCAATATTCA CATTGTGTTC AAGGTCTTTA TTCACCAAGT 2940
 ACGTGTAATG GCCATGATTA ATGTGCCTGT ATGGTTAACC ACTCCAAATA GCTTATATTT 3000
 CATAGTGTCA TTGTTTTTCA ATATAATGTT TAGTATCAAT GGATATGTTA CGACGGTGTT 3060
 ATTTTTCTTG GTCAAATCGT AATAAAATCT CGATAAATGG ATGACTAAGA TTTTGGTAA 3120
 AGTTACAAAA TTTATCGTTT TCACTGTTGT CAATTTTTTG TTCTTGTAAT CACTCGAG 3178

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 816 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPP1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

-- 62 --

```

ATGAAACGTT TCAATGTTTT AAAATATATC AGAACAACAA AAGCAAATAT ACAAACCATC      60
GCAATGCCTT TGACCACAAA ACCTTTATCT TTGAAAATCA ACGCCGCTCT ATTCGATGTT      120
GACGGTACCA TCATCATCTC TCAACCAGCC ATTGCTGCTT TCTGGAGAGA TTTCGGTAAA      180
GACAAGCCTT ACTTCGATGC CGAACACGTT ATTCACATCT CTCACGGTTG GAGAACTTAC      240
GATGCCATTG CCAAGTTCGC TCCAGACTTT GCTGATGAAG AATACGTTAA CAAGCTAGAA      300
GGTGAAATCC CAGAAAAGTA CGGTGAACAC TCCATCGAAG TTCCAGGTGC TGTCAAGTTG      360
TGTAATGCTT TGAACGCCTT GCCAAAGGAA AAATGGGCTG TCGCCACCTC TGGTACCCGT      420
GACATGGCCA AGAAATGGTT CGACATTTTG AAGATCAAGA GACCAGAATA CTTTCATCACC      480
GCCAATGATG TCAAGCAAGG TAAGCCTCAC CCAGAACCAT ACTTAAAGGG TAGAAACGGT      540
TTGGGTTTCC CAATTAATGA ACAAGACCCA TCCAAATCTA AGGTTGTTGT CTTTGAAGAC      600
GCACCAGCTG GTATTGCTGC TGGTAAGGCT GCTGGCTGTA AAATCGTTGG TATTGCTACC      660
ACTTTCGATT TGGACTTCTT GAAGGAAAAG GGTGTGTACA TCATTGTCAA GAACCACGAA      720
TCTATCAGAG TCGGTGAATA CAACGCTGAA ACCGATGAAG TCGAATTGAT CTTTGATGAC      780
TACTTATACG CTAAGGATGA CTTGTTGAAA TGGTAA                                816

```

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 753 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: GPP2
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```

ATGGGATTGA CTAATAAACC TCTATCTTTG AAAGTTAAGC CCGCTTTGTT CGACGTCGAC      60
GGTACCATTA TCATCTCTCA ACCAGCCATT GCTGCATTCT GGAGGGATTT CGGTAAGGAC      120
AAACCTTATT TCGATGCTGA ACACGTTATC CAAGTCTCGC ATGGTTGGAG AACGTTTGAT      180
GCCATTGCTA AGTTCGCTCC AGACTTTGCC AATGAAGAGT ATGTTAACAA ATTAGAAGCT      240
GAAATCCCGG TCAAGTACGG TGAAAAATCC ATTGAAGTCC CAGGTGCAGT TAAGCTGTGC      300
AACGCTTTGA ACGCTCTACC AAAAGAGAAA TGGGCTGTGG CAACTTCCGG TACCCGTGAT      360

```

-- 63 --

ATGGCACAAA AATGGTTCGA GCATCTGGGA ATCAGGAGAC CAAAGTACTT CATTACCGCT 420
 AATGATGTCA AACAGGGTAA GCCTCATCCA GAACCATATC TGAAGGGCAG GAATGGCTTA 480
 GGATATCCGA TCAATGAGCA AGACCCTTCC AAATCTAAGG TAGTAGTATT TGAAGACGCT 540
 CCAGCAGGTA TTGCCGCCGG AAAAGCCGCC GGTGTGAAGA TCATTGGTAT TGCCACTACT 600
 TTCGACTTGG ACTTCCTAAA GGAAAAAGGC TGTGACATCA TTGTCAAAAA CCACGAATCC 660
 ATCAGAGTTG GCGGCTACAA TGCCGAAACA GACGAAGTTG AATTCATTTT TGACGACTAC 720
 TTATATGCTA AGGACGATCT GTTGAAATGG TAA 753

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2520 base pairs.
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GUT1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TGTATTGGCC ACGATAACCA CCCTTTGTAT ACTGTTTTTG TTTTTCACAT GGTAATAAAC 60
 GACTTTTATT AAACAACGTA TGTA AAAACA TAACAAGAAT CTACCCATAC AGGCCATTTC 120
 GTAATTCTTC TCTTCTAATT GGAGTAAAC CATCAATTAA AGGGTGTGGA GTAGCATAGT 180
 GAGGGGCTGA CTGCATTGAC AAAAAAATTG AAAAAAAAAA AGGAAAAGGA AAGGAAAAAA 240
 AGACAGCCAA GACTTTTAGA ACGGATAAGG TGTAATAAAA TGTGGGGGGA TGCCTGTTCT 300
 CGAACCATAT AAAATATACC ATGTGGTTTG AGTTGTGGCC GGAACATAC AAATAGTTAT 360
 ATGTTTCCCT CTCTCTCCG ACTTG TAGTA TTCTCCAAAC GTTACATATT CCGATCAAGC 420
 CAGCGCCTTT AACTAGTTT AAAACAAGAA CAGAGCCGTA TGTCCAAAT AATGGAAGAT 480
 TTACGAAGTG ACTACGTCCC GCTTATCGCC AGTATTGATG TAGGAACGAC CTCATCCAGA 540
 TGCATTCTGT TCAACAGATG GGGCCAGGAC GTTTCAAAAC ACCAAATTGA ATATTCAACT 600
 TCAGCATCGA AGGGCAAGAT TGGGGTGTCT GGCTAAGGA GACCCTCTAC AGCCCCAGCT 660
 CGTGAAACAC CAAACGCCGG TGACATCAAA ACCAGCGGAA AGCCCATCTT TTCTGCAGAA 720
 GGCTATGCCA TTCAAGAAAC CAAATTCCTA AAAATCGAGG AATTGGACTT GGAATTCCAT 780
 AACGAACCCA CGTTGAAGTT CCCCAAACCG GGTGGGTTG AGTGCCATCC GCAGAAATTA 840

-- 64 --

CTGGTGAACG TCGTCCAATG CCTTGCCTCA AGTTTGCTCT CTCTGCAGAC TATCAACAGC 900
GAACGTGTAG CAAACGGTCT CCCACCTTAC AAGGTAATAT GCATGGGTAT AGCAAACATG 960
AGAGAAACCA CAATTCTGTG GTCCCGCCGC ACAGGAAAAC CAATTGTTAA CTACGGTATT 1020
GTTTGGAAACG ACACCAGAAC GATCAAAAATC GTTAGAGACA AATGGCAAAA CACTAGCGTC 1080
GATAGGCAAC TGCAGCTTAG ACAGAAGACT GGATTGCCAT TGCTCTCCAC GTATTTCTCC 1140
TGTTCCAAGC TGCCTGGTT CCTCGACAAT GAGCCTCTGT GTACCAAGGC GTATGAGGAG 1200
AACGACCTGA TGTTCCGCAC TGTGGACACA TGGCTGATTT ACCAATTAAC TAAACAAAAG 1260
GCGTTCGTTT CTGACGTAAC CAACGCTTCC AGAACTGGAT TTATGAACCT CTCCACTTTA 1320
AAGTACGACA ACGAGTTGCT GGAATTTTGG GGTATTGACA AGAACCTGAT TCACATGCCC 1380
GAAATTGTGT CCTCATCTCA ATACTACGGT GACTTTGGCA TTCCTGATTG GATAATGGAA 1440
AAGCTACACG ATTCGCCAAA AACAGTACTG CGAGATCTAG TCAAGAGAAA CCTGCCCATA 1500
CAGGGCTGTC TGGGCGACCA AAGCGCATCC ATGGTGGGGC AACTCGCTTA CAAACCCGGT 1560
GCTGCAAAAT GTACTTATGG TACCGGTTGC TTTTAACTGT ACAATACGGG GACCAAAAAA 1620
TTGATCTCCC AACATGGCGC ACTGACGACT CTAGCATTTT GGTTCCCACA TTTGCAAGAG 1680
TACGGTGGCC AAAAACCAGA ATTGAGCAAG CCACATTTTG CATTAGAGGG TTCCGTCGCT 1740
GTGGCTGGTG CTGTGGTCCA ATGGCTACGT GATAATTTAC GATTGATCGA TAAATCAGAG 1800
GATGTCGGAC CGATTGCATC TACGGTTCCT GATTCTGGTG GCGTAGTTTT CGTCCCCGCA 1860
TTTAGTGGCC TATTCGCTCC CTATTGGGAC CCAGATGCCA GAGCCACCAT AATGGGGATG 1920
TCTCAATTCA CTAATGCCTC CCACATCGCC AGAGCTGCCG TGGAAGGTGT TTGCTTTCAA 1980
GCCAGGGCTA TCTTGAAGGC AATGAGTTCT GACGCGTTTG GTGAAGGTTT CAAAGACAGG 2040
GACTTTTTTAG AGGAAATTTT CGACGTCACA TATGAAAAGT CGCCCCTGTC GGTCTTGGCA 2100
GTGGATGGCG GGATGTCGAG GTCTAATGAA GTCATGCAAA TTCAAGCCGA TATCCTAGGT 2160
CCCTGTGTCA AAGTCAGAAG GTCTCCGACA GCGGAATGTA CCGCATTGGG GGCAGCCATT 2220
GCAGCCAATA TGGCTTTTCAA GGATGTGAAC GAGCGCCCAT TATGGAAGGA CCTACACGAT 2280
GTTAAGAAAT GGGTCTTTTA CAATGGAATG GAGAAAAACG AACAAATATC ACCAGAGGCT 2340
CATCCAAACC TTAAGATATT CAGAAGTGAA TCCGACGATG CTGAAAGGAG AAAGCATTGG 2400
AAGTATTGGG AAGTTGCCGT GGAAAGATCC AAAGGTTGGC TGAAGGACAT AGAAGGTGAA 2460
CACGAACAGG TTCTAGAAAA CTTCCAATAA CAACATAAAT AATTTCTATT AACAATGTAA 2520

-- 65 --

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 391 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPD1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```

Met Ser Ala Ala Ala Asp Arg Leu Asn Leu Thr Ser Gly His Leu Asn
1           5           10           15

Ala Gly Arg Lys Arg Ser Ser Ser Ser Val Ser Leu Lys Ala Ala Glu
20           25           30

Lys Pro Phe Lys Val Thr Val Ile Gly Ser Gly Asn Trp Gly Thr Thr
35           40           45

Ile Ala Lys Val Val Ala Glu Asn Cys Lys Gly Tyr Pro Glu Val Phe
50           55           60

Ala Pro Ile Val Gln Met Trp Val Phe Glu Glu Glu Ile Asn Gly Glu
65           70           75           80

Lys Leu Thr Glu Ile Ile Asn Thr Arg His Gln Asn Val Lys Tyr Leu
85           90           95

Pro Gly Ile Thr Leu Pro Asp Asn Leu Val Ala Asn Pro Asp Leu Ile
100          105          110

Asp Ser Val Lys Asp Val Asp Ile Ile Val Phe Asn Ile Pro His Gln
115          120          125

Phe Leu Pro Arg Ile Cys Ser Gln Leu Lys Gly His Val Asp Ser His
130          135          140

Val Arg Ala Ile Ser Cys Leu Lys Gly Phe Glu Val Gly Ala Lys Gly
145          150          155          160

Val Gln Leu Leu Ser Ser Tyr Ile Thr Glu Glu Leu Gly Ile Gln Cys
165          170          175

Gly Ala Leu Ser Gly Ala Asn Ile Ala Thr Glu Val Ala Gln Glu His
180          185          190

Trp Ser Glu Thr Thr Val Ala Tyr His Ile Pro Lys Asp Phe Arg Gly
195          200          205

Glu Gly Lys Asp Val Asp His Lys Val Leu Lys Ala Leu Phe His Arg

```

-- 66 --

210	215	220
Pro Tyr Phe His Val Ser Val Ile Glu Asp Val Ala Gly Ile Ser Ile		
225	230	235 240
Cys Gly Ala Leu Lys Asn Val Val Ala Leu Gly Cys Gly Phe Val Glu		
	245	250 255
Gly Leu Gly Trp Gly Asn Asn Ala Ser Ala Ala Ile Gln Arg Val Gly		
	260	265 270
Leu Gly Glu Ile Ile Arg Phe Gly Gln Met Phe Phe Pro Glu Ser Arg		
	275	280 285
Glu Glu Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile Thr		
	290	295 300
Thr Cys Ala Gly Gly Arg Asn Val Lys Val Ala Arg Leu Met Ala Thr		
	305	310 315 320
Ser Gly Lys Asp Ala Trp Glu Cys Glu Lys Glu Leu Leu Asn Gly Gln		
	325	330 335
Ser Ala Gln Gly Leu Ile Thr Cys Lys Glu Val His Glu Trp Leu Glu		
	340	345 350
Thr Cys Gly Ser Val Glu Asp Phe Pro Leu Phe Glu Ala Val Tyr Gln		
	355	360 365
Ile Val Tyr Asn Asn Tyr Pro Met Lys Asn Leu Pro Asp Met Ile Glu		
	370	375 380
Glu Leu Asp Leu His Glu Asp		
385	390	

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 384 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPD2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Thr Ala His Thr Asn Ile Lys Gln His Lys His Cys His Glu Asp		
1	5	10 15
His Pro Ile Arg Arg Ser Asp Ser Ala Val Ser Ile Val His Leu Lys		
	20	25 30

-- 67 --

Arg Ala Pro Phe Lys Val Thr Val Ile Gly Ser Gly Asn Trp Gly Thr
 35 40 45
 Thr Ile Ala Lys Val Ile Ala Glu Asn Thr Glu Leu His Ser His Ile
 50 55 60
 Phe Glu Pro Glu Val Arg Met Trp Val Phe Asp Glu Lys Ile Gly Asp
 65 70 75 80
 Glu Asn Leu Thr Asp Ile Ile Asn Thr Arg His Gln Asn Val Lys Tyr
 85 90 95
 Leu Pro Asn Ile Asp Leu Pro His Asn Leu Val Ala Asp Pro Asp Leu
 100 105 110
 Leu His Ser Ile Lys Gly Ala Asp Ile Leu Val Phe Asn Ile Pro His
 115 120 125
 Gln Phe Leu Pro Asn Ile Val Lys Gln Leu Gln Gly His Val Ala Pro
 130 135 140
 His Val Arg Ala Ile Ser Cys Leu Lys Gly Phe Glu Leu Gly Ser Lys
 145 150 155 160
 Gly Val Gln Leu Leu Ser Ser Tyr Val Thr Asp Glu Leu Gly Ile Gln
 165 170 175
 Cys Gly Ala Leu Ser Gly Ala Asn Leu Ala Pro Glu Val Ala Lys Glu
 180 185 190
 His Trp Ser Glu Thr Thr Val Ala Tyr Gln Leu Pro Lys Asp Tyr Gln
 195 200 205
 Gly Asp Gly Lys Asp Val Asp His Lys Ile Leu Lys Leu Leu Phe His
 210 215 220
 Arg Pro Tyr Phe His Val Asn Val Ile Asp Asp Val Ala Gly Ile Ser
 225 230 235 240
 Ile Ala Gly Ala Leu Lys Asn Val Val Ala Leu Ala Cys Gly Phe Val
 245 250 255
 Glu Gly Met Gly Trp Gly Asn Asn Ala Ser Ala Ala Ile Gln Arg Leu
 260 265 270
 Gly Leu Gly Glu Ile Ile Lys Phe Gly Arg Met Phe Phe Pro Glu Ser
 275 280 285
 Lys Val Glu Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile
 290 295 300
 Thr Thr Cys Ser Gly Gly Arg Asn Val Lys Val Ala Thr Tyr Met Ala
 305 310 315 320
 Lys Thr Gly Lys Ser Ala Leu Glu Ala Glu Lys Glu Leu Leu Asn Gly
 325 330 335

-- 68 --

Gln Ser Ala Gln Gly Ile Ile Thr Cys Arg Glu Val His Glu Trp Leu
 340 345 350

Gln Thr Cys Glu Leu Thr Gln Glu Phe Pro Ile Ile Arg Gly Ser Leu
 355 360 365

Pro Asp Ser Leu Gln Gln Arg Pro His Gly Arg Pro Thr Gly Asp Asp
 370 375 380

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 614 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GUT2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Met Thr Arg Ala Thr Trp Cys Asn Ser Pro Pro Pro Leu His Arg Gln
 1 5 10 15

Val Ser Arg Arg Asp Leu Leu Asp Arg Leu Asp Lys Thr His Gln Phe
 20 25 30

Asp Val Leu Ile Ile Gly Gly Gly Ala Thr Gly Thr Gly Cys Ala Leu
 35 40 45

Asp Ala Ala Thr Arg Gly Leu Asn Val Ala Leu Val Glu Lys Gly Asp
 50 55 60

Phe Ala Ser Gly Thr Ser Ser Lys Ser Thr Lys Met Ile His Gly Gly
 65 70 75 80

Val Arg Tyr Leu Glu Lys Ala Phe Trp Glu Phe Ser Lys Ala Gln Leu
 85 90 95

Asp Leu Val Ile Glu Ala Leu Asn Glu Arg Lys His Leu Ile Asn Thr
 100 105 110

Ala Pro His Leu Cys Thr Val Leu Pro Ile Leu Ile Pro Ile Tyr Ser
 115 120 125

Thr Trp Gln Val Pro Tyr Ile Tyr Met Gly Cys Lys Phe Tyr Asp Phe
 130 135 140

Phe Gly Gly Ser Gln Asn Leu Lys Lys Ser Tyr Leu Leu Ser Lys Ser
 145 150 155 160

Ala Thr Val Glu Lys Ala Pro Met Leu Thr Thr Asp Asn Leu Lys Ala

-- 69 --

165					170					175					
Ser	Leu	Val	Tyr	His	Asp	Gly	Ser	Phe	Asn	Asp	Ser	Arg	Leu	Asn	Ala
			180					185					190		
Thr	Leu	Ala	Ile	Thr	Gly	Val	Glu	Asn	Gly	Ala	Thr	Val	Leu	Ile	Tyr
		195					200					205			
Val	Glu	Val	Gln	Lys	Leu	Ile	Lys	Asp	Pro	Thr	Ser	Gly	Lys	Val	Ile
	210					215					220				
Gly	Ala	Glu	Ala	Arg	Asp	Val	Glu	Thr	Asn	Glu	Leu	Val	Arg	Ile	Asn
225				230						235				240	
Ala	Lys	Cys	Val	Val	Asn	Ala	Thr	Gly	Pro	Tyr	Ser	Asp	Ala	Ile	Leu
			245						250					255	
Gln	Met	Asp	Arg	Asn	Pro	Ser	Gly	Leu	Pro	Asp	Ser	Pro	Leu	Asn	Asp
			260					265					270		
Asn	Ser	Lys	Ile	Lys	Ser	Thr	Phe	Asn	Gln	Ile	Ser	Val	Met	Asp	Pro
		275					280					285			
Lys	Met	Val	Ile	Pro	Ser	Ile	Gly	Val	His	Ile	Val	Leu	Pro	Ser	Phe
	290					295					300				
Tyr	Ser	Pro	Lys	Asp	Met	Gly	Leu	Leu	Asp	Val	Arg	Thr	Ser	Asp	Gly
305					310					315					320
Arg	Val	Met	Phe	Phe	Leu	Pro	Trp	Gln	Gly	Lys	Val	Leu	Ala	Gly	Thr
			325						330					335	
Thr	Asp	Ile	Pro	Leu	Lys	Gln	Val	Pro	Glu	Asn	Pro	Met	Pro	Thr	Glu
		340						345					350		
Ala	Asp	Ile	Gln	Asp	Ile	Leu	Lys	Glu	Leu	Gln	His	Tyr	Ile	Glu	Phe
		355					360					365			
Pro	Val	Lys	Arg	Glu	Asp	Val	Leu	Ser	Ala	Trp	Ala	Gly	Val	Arg	Pro
	370					375					380				
Leu	Val	Arg	Asp	Pro	Arg	Thr	Ile	Pro	Ala	Asp	Gly	Lys	Lys	Gly	Ser
385					390					395					400
Ala	Thr	Gln	Gly	Val	Val	Arg	Ser	His	Phe	Leu	Phe	Thr	Ser	Asp	Asn
			405						410					415	
Gly	Leu	Ile	Thr	Ile	Ala	Gly	Gly	Lys	Trp	Thr	Thr	Tyr	Arg	Gln	Met
		420						425					430		
Ala	Glu	Glu	Thr	Val	Asp	Lys	Val	Val	Glu	Val	Gly	Gly	Phe	His	Asn
	435						440					445			
Leu	Lys	Pro	Cys	His	Thr	Arg	Asp	Ile	Lys	Leu	Ala	Gly	Ala	Glu	Glu
	450					455					460				

-- 70 --

Trp Thr Gln Asn Tyr Val Ala Leu Leu Ala Gln Asn Tyr His Leu Ser
 465 470 475 480
 Ser Lys Met Ser Asn Tyr Leu Val Gln Asn Tyr Gly Thr Arg Ser Ser
 485 490 495
 Ile Ile Cys Glu Phe Phe Lys Glu Ser Met Glu Asn Lys Leu Pro Leu
 500 505 510
 Ser Leu Ala Asp Lys Glu Asn Asn Val Ile Tyr Ser Ser Glu Glu Asn
 515 520 525
 Asn Leu Val Asn Phe Asp Thr Phe Arg Tyr Pro Phe Thr Ile Gly Glu
 530 535 540
 Leu Lys Tyr Ser Met Gln Tyr Glu Tyr Cys Arg Thr Pro Leu Asp Phe
 545 550 555 560
 Leu Leu Arg Arg Thr Arg Phe Ala Phe Leu Asp Ala Lys Glu Ala Leu
 565 570 575
 Asn Ala Val His Ala Thr Val Lys Val Met Gly Asp Glu Phe Asn Trp
 580 585 590
 Ser Glu Lys Lys Arg Gln Trp Glu Leu Glu Lys Thr Val Asn Phe Ile
 595 600 605
 Gln Gly Arg Phe Gly Val
 610

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 339 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPSA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Met Asn Gln Arg Asn Ala Ser Met Thr Val Ile Gly Ala Gly Ser Tyr
 1 5 10 15
 Gly Thr Ala Leu Ala Ile Thr Leu Ala Arg Asn Gly His Glu Val Val
 20 25 30
 Leu Trp Gly His Asp Pro Glu His Ile Ala Thr Leu Glu Arg Asp Arg
 35 40 45
 Cys Asn Ala Ala Phe Leu Pro Asp Val Pro Phe Pro Asp Thr Leu His
 50 55 60

-- 71 --

Leu Glu Ser Asp Leu Ala Thr Ala Leu Ala Ala Ser Arg Asn Ile Leu
 65 70 75 80
 Val Val Val Pro Ser His Val Phe Gly Glu Val Leu Arg Gln Ile Lys
 85 90 95
 Pro Leu Met Arg Pro Asp Ala Arg Leu Val Trp Ala Thr Lys Gly Leu
 100 105 110
 Glu Ala Glu Thr Gly Arg Leu Leu Gln Asp Val Ala Arg Glu Ala Leu
 115 120 125
 Gly Asp Gln Ile Pro Leu Ala Val Ile Ser Gly Pro Thr Phe Ala Lys
 130 135 140
 Glu Leu Ala Ala Gly Leu Pro Thr Ala Ile Ser Leu Ala Ser Thr Asp
 145 150 155 160
 Gln Thr Phe Ala Asp Asp Leu Gln Gln Leu Leu His Cys Gly Lys Ser
 165 170 175
 Phe Arg Val Tyr Ser Asn Pro Asp Phe Ile Gly Val Gln Leu Gly Gly
 180 185 190
 Ala Val Lys Asn Val Ile Ala Ile Gly Ala Gly Met Ser Asp Gly Ile
 195 200 205
 Gly Phe Gly Ala Asn Ala Arg Thr Ala Leu Ile Thr Arg Gly Leu Ala
 210 215 220
 Glu Met Ser Arg Leu Gly Ala Ala Leu Gly Ala Asp Pro Ala Thr Phe
 225 230 235 240
 Met Gly Met Ala Gly Leu Gly Asp Leu Val Leu Thr Cys Thr Asp Asn
 245 250 255
 Gln Ser Arg Asn Arg Arg Phe Gly Met Met Leu Gly Gln Gly Met Asp
 260 265 270
 Val Gln Ser Ala Gln Glu Lys Ile Gly Gln Val Val Glu Gly Tyr Arg
 275 280 285
 Asn Thr Lys Glu Val Arg Glu Leu Ala His Arg Phe Gly Val Glu Met
 290 295 300
 Pro Ile Thr Glu Glu Ile Tyr Gln Val Leu Tyr Cys Gly Lys Asn Ala
 305 310 315 320
 Arg Glu Ala Ala Leu Thr Leu Leu Gly Arg Ala Arg Lys Asp Glu Arg
 325 330 335
 Ser Ser His

-- 72 --

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 501 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GLPD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```

Met Glu Thr Lys Asp Leu Ile Val Ile Gly Gly Gly Ile Asn Gly Ala
1           5           10           15
Gly Ile Ala Ala Asp Ala Ala Gly Arg Gly Leu Ser Val Leu Met Leu
          20           25           30
Glu Ala Gln Asp Leu Ala Cys Ala Thr Ser Ser Ala Ser Ser Lys Leu
          35           40           45
Ile His Gly Gly Leu Arg Tyr Leu Glu His Tyr Glu Phe Arg Leu Val
          50           55           60
Ser Glu Ala Leu Ala Glu Arg Glu Val Leu Leu Lys Met Ala Pro His
65           70           75           80
Ile Ala Phe Pro Met Arg Phe Arg Leu Pro His Arg Pro His Leu Arg
          85           90           95
Pro Ala Trp Met Ile Arg Ile Gly Leu Phe Met Tyr Asp His Leu Gly
          100          105          110
Lys Arg Thr Ser Leu Pro Gly Ser Thr Gly Leu Arg Phe Gly Ala Asn
          115          120          125
Ser Val Leu Lys Pro Glu Ile Lys Arg Gly Phe Glu Tyr Ser Asp Cys
          130          135          140
Trp Val Asp Asp Ala Arg Leu Val Leu Ala Asn Ala Gln Met Val Val
          145          150          155          160
Arg Lys Gly Gly Glu Val Leu Thr Arg Thr Arg Ala Thr Ser Ala Arg
          165          170          175
Arg Glu Asn Gly Leu Trp Ile Val Glu Ala Glu Asp Ile Asp Thr Gly
          180          185          190
Lys Lys Tyr Ser Trp Gln Ala Arg Gly Leu Val Asn Ala Thr Gly Pro
          195          200          205
Trp Val Lys Gln Phe Phe Asp Asp Gly Met His Leu Pro Ser Pro Tyr
          210          215          220

```

-- 73 --

Gly Ile Arg Leu Ile Lys Gly Ser His Ile Val Val Pro Arg Val His
 225 230 235 240
 Thr Gln Lys Gln Ala Tyr Ile Leu Gln Asn Glu Asp Lys Arg Ile Val
 245 250 255
 Phe Val Ile Pro Trp Met Asp Glu Phe Ser Ile Ile Gly Thr Thr Asp
 260 265 270
 Val Glu Tyr Lys Gly Asp Pro Lys Ala Val Lys Ile Glu Glu Ser Glu
 275 280 285
 Ile Asn Tyr Leu Leu Asn Val Tyr Asn Thr His Phe Lys Lys Gln Leu
 290 295 300
 Ser Arg Asp Asp Ile Val Trp Thr Tyr Ser Gly Val Arg Pro Leu Cys
 305 310 315 320
 Asp Asp Glu Ser Asp Ser Pro Gln Ala Ile Thr Arg Asp Tyr Thr Leu
 325 330 335
 Asp Ile His Asp Glu Asn Gly Lys Ala Pro Leu Leu Ser Val Phe Gly
 340 345 350
 Gly Lys Leu Thr Thr Tyr Arg Lys Leu Ala Glu His Ala Leu Glu Lys
 355 360 365
 Leu Thr Pro Tyr Tyr Gln Gly Ile Gly Pro Ala Trp Thr Lys Glu-Ser
 370 375 380
 Val Leu Pro Gly Gly Ala Ile Glu Gly Asp Arg Asp Asp Tyr Ala Ala
 385 390 395 400
 Arg Leu Arg Arg Arg Tyr Pro Phe Leu Thr Glu Ser Leu Ala Arg His
 405 410 415
 Tyr Ala Arg Thr Tyr Gly Ser Asn Ser Glu Leu Leu Leu Gly Asn Ala
 420 425 430
 Gly Thr Val Ser Asp Leu Gly Glu Asp Phe Gly His Glu Phe Tyr Glu
 435 440 445
 Ala Glu Leu Lys Tyr Leu Val Asp His Glu Trp Val Arg Arg Ala Asp
 450 455 460
 Asp Ala Leu Trp Arg Arg Thr Lys Gln Gly Met Trp Leu Asn Ala Asp
 465 470 475 480
 Gln Gln Ser Arg Val Ser Gln Trp Leu Val Glu Tyr Thr Gln Gln Arg
 485 490 495
 Leu Ser Leu Ala Ser
 500

(2) INFORMATION FOR SEQ ID NO:16:

-- 74 --

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 542 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GLPABC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```

Met Lys Thr Arg Asp Ser Gln Ser Ser Asp Val Ile Ile Ile Gly Gly
1           5           10           15

Gly Ala Thr Gly Ala Gly Ile Ala Arg Asp Cys Ala Leu Arg Gly Leu
          20           25           30

Arg Val Ile Leu Val Glu Arg His Asp Ile Ala Thr Gly Ala Thr Gly
          35           40           45

Arg Asn His Gly Leu Leu His Ser Gly Ala Arg Tyr Ala Val Thr Asp
          50           55           60

Ala Glu Ser Ala Arg Glu Cys Ile Ser Glu Asn Gln Ile Leu Lys Arg
          65           70           75           80

Ile Ala Arg His Cys Val Glu Pro Thr Asn Gly Leu Phe Ile Thr Leu
          85           90           95

Pro Glu Asp Asp Leu Ser Phe Gln Ala Thr Phe Ile Arg Ala Cys Glu
          100          105          110

Glu Ala Gly Ile Ser Ala Glu Ala Ile Asp Pro Gln Gln Ala Arg Ile
          115          120          125

Ile Glu Pro Ala Val Asn Pro Ala Leu Ile Gly Ala Val Lys Val Pro
          130          135          140

Asp Gly Thr Val Asp Pro Phe Arg Leu Thr Ala Ala Asn Met Leu Asp
          145          150          155          160

Ala Lys Glu His Gly Ala Val Ile Leu Thr Ala His Glu Val Thr Gly
          165          170          175

Leu Ile Arg Glu Gly Ala Thr Val Cys Gly Val Arg Val Arg Asn His
          180          185          190

Leu Thr Gly Glu Thr Gln Ala Leu His Ala Pro Val Val Val Asn Ala
          195          200          205

Ala Gly Ile Trp Gly Gln His Ile Ala Glu Tyr Ala Asp Leu Arg Ile
          210          215          220

Arg Met Phe Pro Ala Lys Gly Ser Leu Leu Ile Met Asp His Arg Ile

```

-- 75 --

225		230		235		240
Asn Gln His Val	Ile Asn Arg Cys Arg	Lys Pro Ser Asp Ala Asp Ile				
	245	250			255	
Leu Val Pro Gly Asp Thr Ile Ser Leu Ile Gly Thr Thr Ser Leu Arg		265			270	
	260					
Ile Asp Tyr Asn Glu Ile Asp Asp Asn Arg Val Thr Ala Glu Glu Val		280			285	
	275					
Asp Ile Leu Leu Arg Glu Gly Glu Lys Leu Ala Pro Val Met Ala Lys		295			300	
	290					
Thr Arg Ile Leu Arg Ala Tyr Ser Gly Val Arg Pro Leu Val Ala Ser		310			315	
						320
Asp Asp Asp Pro Ser Gly Arg Asn Leu Ser Arg Gly Ile Val Leu Leu		325			330	
						335
Asp His Ala Glu Arg Asp Gly Leu Asp Gly Phe Ile Thr Ile Thr Gly		340			345	
						350
Gly Lys Leu Met Thr Tyr Arg Leu Met Ala Glu Trp Ala Thr Asp Ala		355			360	
						365
Val Cys Arg Lys Leu Gly Asn Thr Arg Pro Cys Thr Thr Ala Asp Leu		370			375	
						380
Ala Leu Pro Gly Ser Gln Glu Pro Ala Glu Val Thr Leu Arg Lys Val		385			390	
						395
Ile Ser Leu Pro Ala Pro Leu Arg Gly Ser Ala Val Tyr Arg His Gly		400			405	
						410
Asp Arg Thr Pro Ala Trp Leu Ser Glu Gly Arg Leu His Arg Ser Leu		415			420	
						425
Val Cys Glu Cys Glu Ala Val Thr Ala Gly Glu Val Gln Tyr Ala Val		430			435	
						440
Glu Asn Leu Asn Val Asn Ser Leu Leu Asp Leu Arg Arg Arg Thr Arg		445			450	
						455
Val Gly Met Gly Thr Cys Gln Gly Glu Leu Cys Ala Cys Arg Ala Ala		460			465	
						470
Gly Leu Leu Gln Arg Phe Asn Val Thr Thr Ser Ala Gln Ser Ile Glu		475			480	
						485
Gln Leu Ser Thr Phe Leu Asn Glu Arg Trp Lys Gly Val Gln Pro Ile		490			495	
						500
Ala Trp Gly Asp Ala Leu Arg Glu Ser Glu Phe Thr Arg Trp Val Tyr		505			510	
						515
						520
						525

-- 76 --

Gln Gly Leu Cys Gly Leu Glu Lys Glu Gln Lys Asp Ala Leu
 530 535 540

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 250 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPP2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Met Gly Leu Thr Thr Lys Pro Leu Ser Leu Lys Val Asn Ala Ala Leu
 1 5 10 15

Phe Asp Val Asp Gly Thr Ile Ile Ile Ser Gln Pro Ala Ile Ala Ala
 20 25 30

Phe Trp Arg Asp Phe Gly Lys Asp Lys Pro Tyr Phe Asp Ala Glu His
 35 40 45

Val Ile Gln Val Ser His Gly Trp Arg Thr Phe Asp Ala Ile Ala Lys
 50 55 60

Phe Ala Pro Asp Phe Ala Asn Glu Glu Tyr Val Asn Lys Leu Glu Ala
 65 70 75 80

Glu Ile Pro Val Lys Tyr Gly Glu Lys Ser Ile Glu Val Pro Gly Ala
 85 90 95

Val Lys Leu Cys Asn Ala Leu Asn Ala Leu Pro Lys Glu Lys Trp Ala
 100 105 110

Val Ala Thr Ser Gly Thr Arg Asp Met Ala Gln Lys Trp Phe Glu His
 115 120 125

Leu Gly Ile Arg Arg Pro Lys Tyr Phe Ile Thr Ala Asn Asp Val Lys
 130 135 140

Gln Gly Lys Pro His Pro Glu Pro Tyr Leu Lys Gly Arg Asn Gly Leu
 145 150 155 160

Gly Tyr Pro Ile Asn Glu Gln Asp Pro Ser Lys Ser Lys Val Val Val
 165 170 175

Phe Glu Asp Ala Pro Ala Gly Ile Ala Ala Gly Lys Ala Ala Gly Cys
 180 185 190

Lys Ile Ile Gly Ile Ala Thr Thr Phe Asp Leu Asp Phe Leu Lys Glu
 195 200 205

-- 77 --

Lys Gly Cys Asp Ile Ile Val Lys Asn His Glu Ser Ile Arg Val Gly
210 215 220

Gly Tyr Asn Ala Glu Thr Asp Glu Val Glu Phe Ile Phe Asp Asp Tyr
225 230 235 240

Leu Tyr Ala Lys Asp Asp Leu Leu Lys Trp
245 250

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 709 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

(A) ORGANISM: GUT1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Phe Pro Ser Leu Phe Arg Leu Val Val Phe Ser Lys Arg Tyr Ile
1 5 10 15

Phe Arg Ser Ser Gln Arg Leu Tyr Thr Ser Leu Lys Gln Glu Gln Ser
20 25 30

Arg Met Ser Lys Ile Met Glu Asp Leu Arg Ser Asp Tyr Val Pro Leu
35 40 45

Ile Ala Ser Ile Asp Val Gly Thr Thr Ser Ser Arg Cys Ile Leu Phe
50 55 60

Asn Arg Trp Gly Gln Asp Val Ser Lys His Gln Ile Glu Tyr Ser Thr
65 70 75 80

Ser Ala Ser Lys Gly Lys Ile Gly Val Ser Gly Leu Arg Arg Pro Ser
85 90 95

Thr Ala Pro Ala Arg Glu Thr Pro Asn Ala Gly Asp Ile Lys Thr Ser
100 105 110

Gly Lys Pro Ile Phe Ser Ala Glu Gly Tyr Ala Ile Gln Glu Thr Lys
115 120 125

Phe Leu Lys Ile Glu Glu Leu Asp Leu Asp Phe His Asn Glu Pro Thr
130 135 140

Leu Lys Phe Pro Lys Pro Gly Trp Val Glu Cys His Pro Gln Lys Leu
145 150 155 160

Leu Val Asn Val Val Gln Cys Leu Ala Ser Ser Leu Leu Ser Leu Gln

-- 78 --

165										170					175				
Thr	Ile	Asn	Ser	Glu	Arg	Val	Ala	Asn	Gly	Leu	Pro	Pro	Tyr	Lys	Val				
			180					185					190						
Ile	Cys	Met	Gly	Ile	Ala	Asn	Met	Arg	Glu	Thr	Thr	Ile	Leu	Trp	Ser				
		195					200					205							
Arg	Arg	Thr	Gly	Lys	Pro	Ile	Val	Asn	Tyr	Gly	Ile	Val	Trp	Asn	Asp				
		210				215					220								
Thr	Arg	Thr	Ile	Lys	Ile	Val	Arg	Asp	Lys	Trp	Gln	Asn	Thr	Ser	Val				
225					230					235					240				
Asp	Arg	Gln	Leu	Gln	Leu	Arg	Gln	Lys	Thr	Gly	Leu	Pro	Leu	Leu	Ser				
			245					250						255					
Thr	Tyr	Phe	Ser	Cys	Ser	Lys	Leu	Arg	Trp	Phe	Leu	Asp	Asn	Glu	Pro				
			260					265					270						
Leu	Cys	Thr	Lys	Ala	Tyr	Glu	Glu	Asn	Asp	Leu	Met	Phe	Gly	Thr	Val				
		275					280					285							
Asp	Thr	Trp	Leu	Ile	Tyr	Gln	Leu	Thr	Lys	Gln	Lys	Ala	Phe	Val	Ser				
		290				295					300								
Asp	Val	Thr	Asn	Ala	Ser	Arg	Thr	Gly	Phe	Met	Asn	Leu	Ser	Thr	Leu				
305					310					315					320				
Lys	Tyr	Asp	Asn	Glu	Leu	Leu	Glu	Phe	Trp	Gly	Ile	Asp	Lys	Asn	Leu				
			325					330						335					
Ile	His	Met	Pro	Glu	Ile	Val	Ser	Ser	Ser	Gln	Tyr	Tyr	Gly	Asp	Phe				
		340						345					350						
Gly	Ile	Pro	Asp	Trp	Ile	Met	Glu	Lys	Leu	His	Asp	Ser	Pro	Lys	Thr				
		355					360				365								
Val	Leu	Arg	Asp	Leu	Val	Lys	Arg	Asn	Leu	Pro	Ile	Gln	Gly	Cys	Leu				
		370				375					380								
Gly	Asp	Gln	Ser	Ala	Ser	Met	Val	Gly	Gln	Leu	Ala	Tyr	Lys	Pro	Gly				
385					390					395					400				
Ala	Ala	Lys	Cys	Thr	Tyr	Gly	Thr	Gly	Cys	Phe	Leu	Leu	Tyr	Asn	Thr				
			405						410					415					
Gly	Thr	Lys	Lys	Leu	Ile	Ser	Gln	His	Gly	Ala	Leu	Thr	Thr	Leu	Ala				
			420					425					430						
Phe	Trp	Phe	Pro	His	Leu	Gln	Glu	Tyr	Gly	Gly	Gln	Lys	Pro	Glu	Leu				
		435					440					445							
Ser	Lys	Pro	His	Phe	Ala	Leu	Glu	Gly	Ser	Val	Ala	Val	Ala	Gly	Ala				
		450				455					460								

-- 79 --

Val Val Gln Trp Leu Arg Asp Asn Leu Arg Leu Ile Asp Lys Ser Glu
 465 470 475 480
 Asp Val Gly Pro Ile Ala Ser Thr Val Pro Asp Ser Gly Gly Val Val
 485 490 495
 Phe Val Pro Ala Phe Ser Gly Leu Phe Ala Pro Tyr Trp Asp Pro Asp
 500 505 510
 Ala Arg Ala Thr Ile Met Gly Met Ser Gln Phe Thr Thr Ala Ser His
 515 520 525
 Ile Ala Arg Ala Ala Val Glu Gly Val Cys Phe Gln Ala Arg Ala Ile
 530 535 540
 Leu Lys Ala Met Ser Ser Asp Ala Phe Gly Glu Gly Ser Lys Asp Arg
 545 550 555 560
 Asp Phe Leu Glu Glu Ile Ser Asp Val Thr Tyr Glu Lys Ser Pro Leu
 565 570 575
 Ser Val Leu Ala Val Asp Gly Gly Met Ser Arg Ser Asn Glu Val Met
 580 585 590
 Gln Ile Gln Ala Asp Ile Leu Gly Pro Cys Val Lys Val Arg Arg Ser
 595 600 605
 Pro Thr Ala Glu Cys Thr Ala Leu Gly Ala Ala Ile Ala Ala Asn Met
 610 615 620
 Ala Phe Lys Asp Val Asn Glu Arg Pro Leu Trp Lys Asp Leu His Asp
 625 630 635 640
 Val Lys Lys Trp Val Phe Tyr Asn Gly Met Glu Lys Asn Glu Gln Ile
 645 650 655
 Ser Pro Glu Ala His Pro Asn Leu Lys Ile Phe Arg Ser Glu Ser Asp
 660 665 670
 Asp Ala Glu Arg Arg Lys His Trp Lys Tyr Trp Glu Val Ala Val Glu
 675 680 685
 Arg Ser Lys Gly Trp Leu Lys Asp Ile Glu Gly Glu His Glu Gln Val
 690 695 700
 Leu Glu Asn Phe Gln
 705

(2) INFORMATION FOR SEQ ID NO:19:

(i) --SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12145 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

-- 80 --

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: PHK28-26

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GTCGACCACC	ACGGTGGTGA	CTTTAATGCC	GCTCTCATGC	AGCAGCTCGG	TGGCGGTCTC	60
AAAATTCAGG	ATGTCGCCGG	TATAGTTTTT	GATAATCAGC	AAGACGCCTT	CGCCGCCGTC	120
AATTTGCATC	GCGCATTCAA	ACATTTTGTC	CGGCGTCGGC	GAGGTGAATA	TTTCCCCCGG	180
ACAGGCGCCG	GAGAGCATGC	CCTGGCCGAT	ATAGCCGCAG	TGCATCGGTT	CATGTCCGCT	240
GCCGCCGCCG	GAGAGCAGGG	CCACCTTGCC	AGCCACCGGC	GCGTCGGTGC	GGGTCACATA	300
CAGCGGGTCC	TGATGCAGGG	TCAGCTGCGG	ATGGGCTTTA	GCCAGCCCCT	GTAATTGTTC	360
ATTCAGTACA	TCTTCAACAC	GGTTAATCAG	CTTTTTCATT	ATTCAGTGCT	CCGTTGGAGA	420
AGGTTTCGATG	CCGCCTCTCT	GCTGGCGGAG	GCGGTCATCG	CGTAGGGGTA	TCGTCTGACG	480
GTGGAGCGTG	CCTGGCGATA	TGATGATTCT	GGCTGAGCGG	ACGAAAAAAAA	GAATGCCCCG	540
ACGATCGGGT	TTCATTACGA	AACATTGCTT	CCTGATTTTG	TTTCTTTATG	GAACGTTTTT	600
GCTGAGGATA	TGGTGAAAAT	GCGAGCTGGC	GCGCTTTTTT	TCTTCTGCCA	TAAGCGGCGG	660
TCAGGATAGC	CGGCGAAGCG	GGTGGGAAAA	AATTTTTTGC	TGATTTTCTG	CCGACTGCGG	720
GAGAAAAGGC	GGTCAAACAC	GGAGGATTGT	AAGGGCATT	TGCGGCAAAG	GAGCGGATCG	780
GGATCGCAAT	CCTGACAGAG	ACTAGGGTTT	TTTGTTCCAA	TATGGAACGT	AAAAAATTAA	840
CCTGTGTTTC	ATATCAGAAC	AAAAGGCGA	AAGATTTTTT	TGTTCCCTGC	CGGCCCTACA	900
GTGATCGCAC	TGCTCCGGTA	CGCTCCGTTC	AGGCCGCGCT	TCACTGGCCG	GCGCGGATAA	960
CGCCAGGGCT	CATCATGTCT	ACATGCGCAC	TTATTTGAGG	GTGAAAGGAA	TGCTAAAAGT	1020
TATTCAATCT	CCAGCCAAAT	ATCTTCAGGG	TCCTGATGCT	GCTGTTCTGT	TCGGTCAATA	1080
TGCCAAAAAC	CTGGCGGAGA	GCTTCTTCGT	CATCGCTGAC	GATTTTCGTAA	TGAAGCTGGC	1140
GGGAGAGAAA	GTGGTGAATG	GCCTGCAGAG	CCACGATATT	CGCTGCCATG	CGGAACGGTT	1200
TAACGGCGAA	TGCAGCCATG	CGGAAATCAA	CCGTCTGATG	GCGATTTTGC	AAAAACAGGG	1260
CTGCCGCGGC	GTGGTCGGGA	TCGGCGGTGG	TAAAACCCTC	GATACGCGCA	AGGCGATCGG	1320
TTACTACCAG	AAGCTGCCGG	TGGTGGTGAT	CCCGACCATC	GCCTCGACCG	ATGCGCCAAC	1380
CAGCGCGCTG	TCGGTGATCT	ACACCGAAGC	GGGCGAGTTT	GAAGAGTATC	TGATCTATCC	1440
GAAAAACCCG	GATATGGTGG	TGATGGACAC	GGCGATTATC	GCCAAAGCGC	CGGTACGCCT	1500

-- 81 --

GCTGGTCTCC GGCATGGGCG ATGCGCTCTC CACCTGGTTC GAGGCCAAAG CTTGCTACGA 1560
TGCGCGCGCC ACCAGCATGG CCGGAGGACA GTCCACCGAG GCGGCGCTGA GCCTCGCCCG 1620
CCTGTGCTAT GATACGCTGC TGGCGGAGGG CGAAAAGGCC CGTCTGGCGG CGCAGGCCGG 1680
GGTAGTGACC GAAGCGCTGG AGCGCATCAT CGAGGCGAAC ACTTACCTCA GCGGCATTGG 1740
CTTTGAAAGC AGTGGCCTGG CCGCTGCCCA TGCAATCCAC AACGGTTTCA CCATTCTTGA 1800
AGAGTGCCAT CACCTGTATC ACGGTGAGAA AGTGGCCTTC GGTACCCTGG CGCAGCTGGT 1860
GCTGCAGAAC AGCCCGATGG ACGAGATTGA AACGGTGCAG GGCTTCTGCC AGCGCGTCGG 1920
CCTGCCGGTG ACGCTCGCGC AGATGGGCGT CAAAGAGGGG ATCGACGAGA AAATCGCCGC 1980
GGTGGCGAAA GCTACCTGCG CGGAAGGGGA AACCATCCAT AATATGCCGT TTGCGGTGAC 2040
CCCGGAGAGC GTCCATGCCG CTATCCTCAC CGCCGATCTG TTAGGCCAGC AGTGGCTGGC 2100
GCGTTAATTC GCGGTGGCTA AACCGCTGGC CCAGGTCAGC GGTTTTCTT TCTCCCCTCC 2160
GGCAGTCGCT GCCGGAGGGG TTCTCTATGG TACAACGCGG AAAAGGATAT GACTGTTTCA 2220
ACTCAGGATA CCGGGAAGGC GGTCTCTTCC GTCATTGCCC AGTCATGGCA CCGCTGCAGC 2280
AAGTTTATGC AGCGCGAAAC CTGGCAAACG CCGCACCAGG CCCAGGGCCT GACCTTCGAC 2340
TCCATCTGTC GGCCTAAAAC CGCGCTGCTC ACCATCGGCC AGGCGGCGCT GGAAGACGCC 2400
TGGGAGTTTA TGGACGGCCG CCCCTGCGCG CTGTTTATTC TTGATGAGTC CGCCTGCATC 2460
CTGAGCCGTT GCGGCGAGCC GCAAACCCTG GCCCAGCTGG CTGCCCTGGG ATTTTCGCGAC 2520
GGCAGCTATT GTGCGGAGAG CATTATCGGC ACCTGCGCGC TGTCGCTGGC CGCGATGCAG 2580
GGCCAGCCGA TCAACACCGC CGGCGATCGG CATTTTAAGC AGGCGCTACA GCCATGGAGT 2640
TTTTGCTCGA CGCCGGTGTT TGATAACCAC GGGCGGCTGT TCGGCTCTAT CTCGCTTTGC 2700
TGTCTGGTCG AGCACCAGTC CAGCGCCGAC CTCTCCCTGA CGCTGGCCAT CGCCCGCGAG 2760
GTGGGTAACCT CCCTGCTTAC CGACAGCCTG CTGGCGGAAT CCAACCGTCA CCTCAATCAG 2820
ATGTACGGCC TGCTGGAGAG CATGGACGAT GGGGTGATGG CGTGGAACGA ACAGGGCGTG 2880
CTGCAGTTTC TCAATGTTCA GGCGGCGAGA CTGCTGCATC TTGATGCTCA GGCCAGCCAG 2940
GGGAAAAATA TCGCCGATCT GGTGACCCTC CCGGCGCTGC TCGCGCGCGC CATCAAACAC 3000
GCCCCGCGCC TGAATCACGT CGAAGTCACC TTTGAAAGTC AGCATCAGTT TGTCGATGCG 3060
GTGATCACCT TAAAACCGAT TGTCGAGGCG CAAGGCAACA GTTTTATTCT GCTGCTGCAT 3120
CCGGTGGAGC AGATGCGGCA GCTGATGACC AGCCAGCTCG GTAAAGTCAG CCACACCTTT 3180

-- 82 --

GAGCAGATGT CTGCCGACGA TCCGGAAACC CGACGCCTGA TCCACTTTGG CCGCCAGGCG 3240
GCGCGCGGCG GCTTCCCGGT GCTACTGTGC GCGAAGAGG GGGTCGGGAA AGAGCTGCTG 3300
AGCCAGGCTA TTCACAATGA AAGCGAACGG GCGGGCGGCC CCTACATCTC CGTCAACTGC 3360
CAGCTATATG CCGACAGCGT GCTGGGCCAG GACTTTATGG GCAGCGCCCC TACCGACGAT 3420
GAAAATGGTC GCCTGAGCCG CCTTGAGCTG GCCAACGGCG GCACCCTGTT TCTGGAAAAG 3480
ATCGAGTATC TGGCGCCGGA GCTGCAGTCG GCTCTGCTGC AGGTGATTAA GCAGGGCGTG 3540
CTCACCCGCC TCGACGCCCC GCGCCTGATC CCGGTGGATG TGAAGGTGAT TGCCACCACC 3600
ACCGTCGATC TGGCCAATCT GGTGGAACAG AACCGCTTTA GCCGCCAGCT GTACTATGCG 3660
CTGCACTCCT TTGAGATCGT CATCCCGCCG CTGCGCGCCC GACGCAACAG TATTCCGTCG 3720
CTGGTGATA ACCGGTTGAA GAGCCTGGAG AAGCGTTTCT CTTCGCGACT GAAAGTGGAC 3780
GATGACGCGC TGGCACAGCT GGTGGCCTAC TCGTGGCCGG GGAATGATTT TGAGCTCAAC 3840
AGCGTCATTG AGAATATCGC CATCAGCAGC GACAACGGCC ACATTCGCCT GAGTAATCTG 3900
CCGGAATATC TCTTTTCCGA GCGGCCGGGC GGGGATAGCG CGTCATCGCT GCTGCCGGCC 3960
AGCCTGACTT TTAGCGCCAT CGAAAAGGAA GCTATTATTC ACGCCGCCCG GGTGACCAGC 4020
GGGCGGGTGC AGGAGATGTC GCAGCTGCTC AATATCGGCC GCACCACCCT GTGGCGCAAA 4080
ATGAAGCAGT ACGATATTGA CGCCAGCCAG TTCAAGCGCA AGCATCAGGC CTAGTCTCTT 4140
CGATTGCGC CATGGAGAAC AGGGCATCCG ACAGGCGATT GCTGTAGCGT TTGAGCGCGT 4200
CGCGCAGCGG ATGCGCGCGG TCCATGGCCG TCAGCAGGCG TTCGAGCCGA CGGGACTGGG 4260
TGCGCGCCAC GTGCAGCTGG GCAGAGGCGA GATTCTCCC CGGGATCACG AACTGTTTTA 4320
ACGGGCCGCT CTCGGCCATA TTGCGGTCGA TAAGCCGCTC CAGGGCGGTG ATCTCCTCTT 4380
CGCCGATCGT CTGGCTCAGG CGGGTCAGGC CCCGCGCATC GCTGGCCAGT TCAGCCCCCA 4440
GCACGAACAG CGTCTGCTGA ATATGGTGCA GGCTTTCCCG CAGCCCGGCG TCGCGGGTCG 4500
TGGCGTAGCA GACGCCCAGC TGGGATATCA GTTCATCGAC GGTGCCGTAG GCCTCGACGC 4560
GAATATGGTC TTTCTCGATG CGGCTGCCGC CGTACAGGGC GGTGGTGCCT TTATCCCCGG 4620
TGCGGGTATA GATACGATAC ATTCAGTTTC TCTCACTTAA CGGCAGGACT TTAACCAGCT 4680
GCCCCGCGTT GGCGCCGAGC GTACGCAGTT GATCGTCGCT ATCGGTGACG TGTCCGGTAG 4740
CCAGCGGCGC GTCCGCCGGC AGCTGGGCAT GAGTGAGGGC TATCTGCCG GACGCGCTGA 4800
GCCCCGATACC CACCCGCAGG GGCGAGCTTC TGGCCGCCAG GCGGCCCAGC GCAGCGGCGT 4860

-- 83 --

CACCGCCTCC GTCATAGGTT ATGGTCTGGC AGGGGACCCC CTGCTCCTCC AGCCCCCAGC 4920
ACAGCTCATT GATGGCGCCG GCATGGTGCC CGCGCGGATC GTAAACAGG CGTACGCCTG 4980
GCGGTGAAAG CGACATGACG GTCCCCTCGT TAACACTCAG AATGCCTGGC GGAAAATCGC 5040
GGCAATCTCC TGCTCGTTGC CTTTACGCGG GTTCGAGAAC GCATTGCCGT CTTTATAGAGC 5100
CATCTCCGCC ATGTAGGGGA AGTCGGCCTC TTTTACCCCC AGATCGCGCA GATGCTGCGG 5160
AATACCGATA TCCATCGACA GACGCGTGAT AGCGGCGATG GCTTTTTCCG CCGCGTCGAG 5220
AGTGGACAGT CCGGTGATAT TTTCGCCCAT CAGTTCAGCG ATATCGGCGA ATTTCTCCGG 5280
GTTGGCGATC AGGTTGTAGC GCGCCACATG CGGCAGCAGG ACAGCGTTGG CCACGCCGTG 5340
CGGCATGTCT TACAGGCCGC CCAGCTGGTG CGCCATGGCG TGCACGTAGC CGAGGTTGGC 5400
GTTATTGAAA GCCATCCCGG CCAGCAGAGA AGCATAGGCC ATGTTTTCCC GCGCCTGCAG 5460
ATTGCTGCCG AGGGCCACGG CCTGGCGCAG GTTGCGGGCG ATGAGGCGGA TCGCCTGCAT 5520
GGCGGCGGCG TCCGTCACCG GGTAGCGTC TTTGGAGATA TAGGCCTCTA CGGCGTGGGT 5580
CAGGGCATCC ATCCCGGTCG CCGCGGTCAG GCGGCGCGGT TTACCGATCA TCAGCAGTGG 5640
ATCGTTGATA GAGACCGACG GCAGTTTGCG CCAGCTGACG ATCACAACT TCACTTTGGT 5700
TTCGGTGTTG GTCAGGACGC AGTGGCGGGT GACCTCGCTG GCGGTGCCGG CGGTGGTATT 5760
GACCGCGACG ATAGGCGGCA GCGGGTTGGT CAGGGTCTCG ATTCCGGCAT ACTGGTACAG 5820
ATCGCCCTCA TGGGTGGCGG CGATGCCGAT GCCTTTGCCG CAATCGTGCG GGCTGCCGCC 5880
GCCCACGGTG ACGATGATGT CGCACTGTTC GCGGCGAAAC ACGGCGAGGC CGTCGCGCAC 5940
GTTGGTGTCT TTCGGGTTCG GCTCGACGCC GTCAAAGATC GCCACCTCGA TCCCGGCCTC 6000
CCGCAGATAA TGCAGGGTTT TGTCCACCGC GCCATCTTTA ATTGCCCCGA GGCCTTTGTC 6060
GGTGACCAGC AGGGCTTTTT TCCCCCCCAG CAGCTGGCAG CGTTCGCCGA CTACGGAAAT 6120
GGCGTTGGGG CCAAAAAAGT TAACGTTTGG CACCAGATAA TCAAACATAC GATAGCTCAT 6180
AATATACCTT CTCGCTTCAG GTTATAATGC GGAAAAACAA TCCAGGGCGC ACTGGGCTAA 6240
TAATTGATCC TGCTCGACCG TACCGCCGCT AACGCCGACG GCGCCAATTA CCTGCTCATT 6300
AAAAATAACT GGCAGGCCGC CGCCAAAAAT AATAATTCGC TGTGGTTGG TTAGCTGCAG 6360
ACCGTACAGA GATTGTCCTG GCTGGACCGC TGACGTAATT TCATGGGTAC CTTGCTTCAG 6420
GCTGCAGGCG CTCCAGGCTT TATTCAGGGA AATATCGCAG CTGGAGACGA AGGCCTCGTC 6480
CATCCGCTGG ATAAGCAGCG TGTTGCCTCC GCGGTCAACT ACGGAAAACA CCACCGCCAC 6540

-- 84 --

GTTGATCTCA GTGGCTTTTT TTTCCACCGC CGCCGCCATT TGCTGGGCGG CGGCCAGGGT 6600
GATTGTCTGA ACTTGTTGGC TCTTGTTTTCAT CATTCTCTCC CGCACCAGGA TAACGCTGGC 6660
GCGAATAGTC AGTAGGGGGC GATAGTAAAA AACTATTACC ATTCGGTTGG CTTGCTTTAT 6720
TTTTGTGAGC GTTATTTTGT CGCCCGCCAT GATTTAGTCA ATAGGGTTAA AATAGCGTCG 6780
GAAAAACGTA ATTAAGGGCG TTTTTTATTA ATTGATTTAT ATCATTGCGG GCGATCACAT 6840
TTTTTATTTT TGCCGCCGGA GTAAAGTTTC ATAGTGAAAC TGTCGGTAGA TTTCGTGTGC 6900
CAAATTGAAA CGAAATTAAA TTTATTTTTT TCACCACTGG CTCATTTAAA GTTCCGCTAT 6960
TGCCGGTAAT GGCCGGGCGG CAACGACGCT GGCCCGGCGT ATTCGCTACC GTCTGCGGAT 7020
TTCACCTTTT GAGCCGATGA ACAATGAAAA GATCAAAACG ATTTGCAGTA CTGGCCCAGC 7080
GCCCCGTCAA TCAGGACGGG CTGATTGGCG AGTGGCCTGA AGAGGGGCTG ATCGCCATGG 7140
ACAGCCCCTT TGACCCGGTC TCTTCAGTAA AAGTGGACAA CGGTCTGATC GTCGAACTGG 7200
ACGGCAAACG CCGGGACCAG TTTGACATGA TCGACCGATT TATCGCCGAT TACGCGATCA 7260
ACGTTGAGCG CACAGAGCAG GCAATGCGCC TGGAGGCGGT GGAAATAGCC CGTATGCTGG 7320
TGGATATTCA CGTCAGCCGG GAGGAGATCA TTGCCATCAC TACCGCCATC ACGCCGGCCA 7380
AAGCGGTCTGA GGTGATGGCG CAGATGAACG TGGTGGAGAT GATGATGGCG CTGCAGAAGA 7440
TGCGTGCCCG CCGGACCCCC TCCAACCAGT GCCACGTCAC CAATCTCAA GATAATCCGG 7500
TGCAGATTGC CGCTGACGCC GCCGAGGCCG GGATCCGCGG CTTCTCAGAA CAGGAGACCA 7560
CGGTGCGTAT CGCGCGCTAC GCGCCGTTTA ACGCCCTGGC GCTGTTGGTC GGTTCGCAGT 7620
GCGGCCGCC CCGCGTGTTG ACGCAGTGCT CCGTGGAAGA GGCCACCGAG CTGGAGCTGG 7680
GCATGCGTGG CTTAACCAGC TACGCCGAGA CCGTGTCGGT CTACGGCACC GAAGCGGTAT 7740
TTACCGACGG CGATGATACG CCGTGGTCAA AGGCGTTTCT CGCCTCGGCC TACGCCTCCC 7800
GCGGGTTGAA AATGCGCTAC ACCTCCGGCA CCGGATCCGA AGCGCTGATG GGCTATTCGG 7860
AGAGCAAGTC GATGCTCTAC CTCGAATCGC GCTGCATCTT CATTACTAAA GGCGCCGGGG 7920
TTCAGGGACT GCAAAACGGC GCGGTGAGCT GTATCGGCAT GACCGGCGCT GTGCCGTCGG 7980
GCATTCGGGC GGTGCTGGCG GAAAACCTGA TCGCCTCTAT GCTCGACCTC GAAGTGGCGT 8040
CCGCCAACGA CCAGACTTTC TCCCACTCGG ATATTGCGCG CACCGCGCGC ACCCTGATGC 8100
AGATGCTGCC GGGCACCGAC TTTATTTTCT CCGGCTACAG CGCGGTGCCG AACTACGACA 8160
ACATGTTTCG CCGCTCGAAC TTCGATGCGG AAGATTTTGA TGATTACAAC ATCCTGCAGC 8220

-- 85 --

GTGACCTGAT GGTGACGGC GGCCTGCGTC CGGTGACCGA GGCGGAAACC ATTGCCATTC 8280
GCCAGAAAGC GGCGCGGGCG ATCCAGGCGG TTTTCCGCGA GCTGGGGCTG CCGCCAATCG 8340
CCGACGAGGA GGTGGAGGCC GCCACCTACG CGCACGGCAG CAACGAGATG CCGCCGCGTA 8400
ACGTGGTGA GATCTGAGT GCGGTGGAAG AGATGATGAA GCGCAACATC ACCGGCCTCG 8460
ATATTGTCGG CGCGCTGAGC CGCAGCGGCT TTGAGGATAT CGCCAGCAAT ATTCTCAATA 8520
TGCTGCGCCA GCGGGTCACC GGCATTACC TGCAGACCTC GGCCATTCTC GATCGGCAGT 8580
TCGAGGTGGT GAGTGCGGTC AACGACATCA ATGACTATCA GGGGCCGGGC ACCGGCTATC 8640
GCATCTCTGC CGAACGCTGG GCGGAGATCA AAAATATTCC GGGCGTGGTT CAGCCCGACA 8700
CCATTGAATA AGGCGGTATT CCTGTGCAAC AGACAACCCA AATTCAGCCC TCTTTTACCC 8760
TGAAAACCCG CGAGGGCGGG GTAGCTTCTG CCGATGAACG CGCCGATGAA GTGGTGATCG 8820
GCGTCGGCCC TGCCTTCGAT AAACACCAGC ATCACACTCT GATCGATATG CCCCATGGCG 8880
CGATCCTCAA AGAGCTGATT GCCGGGGTGG AAGAAGAGGG GCTTCACGCC CGGGTGGTGC 8940
GCATTCTGCG CACGTCCGAC GTCTCCTTTA TGGCCTGGGA TGCGGCCAAC CTGAGCGGCT 9000
CGGGGATCGG CATCGGTATC CAGTCGAAGG GGACCACGGT CATCCATCAG CGCGATCTGC 9060
TGCCGCTCAG CAACCTGGAG CTGTTCTCCC AGGCGCCGCT GCTGACGCTG GAGACCTACC 9120
GGCAGATTGG CAAAACGCT GCGCGCTATG CGCGCAAAGA GTCACCTTCG CCGGTGCCGG 9180
TGGTGAACGA TCAGATGGTG CGGCCGAAAT TTATGGCCAA AGCCGCGCTA TTTCATATCA 9240
AAGAGACCAA ACATGTGGTG CAGGACGCCG AGCCCGTCAC CCTGCACATC GACTTAGTAA 9300
GGGAGTGACC ATGAGCGAGA AAACCATGCG CGTGCAGGAT TATCCGTTAG CCACCCGCTG 9360
CCCGGAGCAT ATCCTGACGC CTACCGGCAA ACCATTGACC GATATTACCC TCGAGAAGGT 9420
GCTCTCTGGC GAGGTGGGCC CGCAGGATGT GCGGATCTCC CGCCAGACCC TTGAGTACCA 9480
GGCGCAGATT GCCGAGCAGA TGCAGCGCCA TGCGGTGGCG CGCAATTTCC GCCGCGCGGC 9540
GGAGCTTATC GCCATTCTTG ACGAGCGCAT TCTGGCTATC TATAACGCGC TGCGCCCGTT 9600
CCGCTCCTCG CAGGCGGAGC TGCTGGCGAT CGCCGACGAG CTGGAGCACA CCTGGCATGC 9660
GACAGTGAAT GCCGCCTTTG TCCGGGAGTC GGCGGAAGTG TATCAGCAGC GGCATAAGCT 9720
GCGTAAAGGA AGCTAAGCGG AGGTCAGCAT GCCGTTAATA GCCGGGATTG ATATCGGCAA 9780
CGCCACCACC GAGGTGGCGC TGGCGTCCGA CTACCCGCAG GCGAGGGCGT TTGTTGCCAG 9840
CGGGATCGTC GCGACGACGG GCATGAAAGG GACGCGGGAC AATATCGCCG GGACCCTCGC 9900

-- 86 --

CGCGCTGGAG CAGGCCCTGG CGAAAACACC GTGGTCGATG AGCGATGTCT CTCGCATCTA 9960
TCTTAACGAA GCCGCGCCGG TGATTGGCGA TGTGGCGATG GAGACCATCA CCGAGACCAT 10020
TATCACCGAA TCGACCATGA TCGGTCATAA CCCGCAGACG CCGGGCGGGG TGGGCGTTGG 10080
CGTGGGGACG ACTATCGCCC TCGGGCGGGT GGCACGCTG CCGGCGGCGC AGTATGCCGA 10140
GGGGTGGATC GTACTGATTG ACGACGCCGT CGATTTCCTT GACGCCGTGT GGTGGCTCAA 10200
TGAGGCGCTC GACCGGGGGA TCAACGTGGT GCGGCGCATC CTCAAAAGG ACGACGGCGT 10260
GCTGGTGAAC AACC GCCTGC GTAAAACCCT GCCGGTGGTG GATGAAGTGA CGCTGCTGGA 10320
GCAGGTCCCC GAGGGGGTAA TGGCGGCGGT GGAAGTGGCC GCGCCGGGCC AGGTGGTGCG 10380
GATCCTGTCT AATCCCTACG GGATCGCCAC CTTCTTCGGG CTAAGCCCGG AAGAGACCCA 10440
GGCCATCGTC CCCATCGCCC GCGCCCTGAT TGGCAACCGT TCCGCGGTGG TGCTCAAGAC 10500
CCCGCAGGGG GATGTGCAGT CGCGGGTGAT CCCGGCGGGC AACCTCTACA TTAGCGGCGA 10560
AAAGCGCCGC GGAGAGGCCG ATGTCGCCGA GGGCGCGGAA GCCATCATGC AGGCGATGAG 10620
CGCCTGCGCT CCGGTACGCG ACATCCGCGG CGAACC GGCG ACCCACGCCG GCGGCATGCT 10680
TGAGCGGGTG CGCAAGGTAA TGGCGTCCCT GACCGGCCAT GAGATGAGCG CGATATACAT 10740
CCAGGATCTG CTGGCGGTGG ATACGTTTAT TCCGCGCAAG GTGCAGGGCG GGATGGCCGG 10800
CGAGTGCGCC ATGGAGAATG CCGTCGGGAT GCGGCGCATG GTGAAAGCGG ATCGTCTGCA 10860
AATGCAGGTT ATCGCCCGCG AACTGAGCGC CCGACTGCAG ACCGAGGTGG TGGTGGGCGG 10920
CGTGGAGGCC AACATGGCCA TCGCCGGGGC GTTAACCACT CCCGGCTGTG CGGCGCCGCT 10980
GGCGATCCTC GACCTCGGCG CCGGCTCGAC GGATGCGGCG ATCGTCAACG CGGAGGGGCA 11040
GATAACGGCG GTCCATCTCG CCGGGGCGGG GAATATGGTC AGCCTGTTGA TTAAAACCGA 11100
GCTGGGCCTC GAGGATCTTT CGCTGGCGGA AGCGATAAAA AAATACCCGC TGGCCAAAGT 11160
GGAAAGCCTG TTCAGTATTC GTCACGAGAA TGGCGCGGTG GAGTTCTTTC GGGAAGCCCT 11220
CAGCCCGGCG GTGTTCGCCA AAGTGGTGTA CATCAAGGAG GGCGAACTGG TGCCGATCGA 11280
TAACGCCAGC CCGCTGGAAA AAATTCGTCT CGTGCGCCGG CAGGCGAAAG AGAAAGTGTT 11340
TGTCACCAAC TGCTGCGCG CGCTGCGCCA GGTCTCACC GCGGTTCCA TTCGCGATAT 11400
CGCCTTTGTG GTGCTGGTGG GCGGCTCATC GCTGGACTTT GAGATCCCGC AGCTTATCAC 11460
GGAAGCCTTG TCGCACTATG GCGTGGTCGC CGGGCAGGGC AATATTCGGG GAACAGAAGG 11520
GCCGCGCAAT GCGGTCGCCA CCGGGCTGCT ACTGGCCGGT CAGGCGAATT AAACGGGCGC 11580

-- 87 --

TCGCGCCAGC CTCTCTCTTT AACGTGCTAT TTCAGGATGC CGATAATGAA CCAGACTTCT 11640
 ACCTTAACCG GGCAGTGCCT GGCCGAGTTT CTTGGCACCG GATTGCTCAT TTTCTTCGGC 11700
 GCGGGCTGCG TCGCTGCGCT GCGGGTCGCC GGGGCCAGCT TTGGTCAGTG GGAGATCAGT 11760
 ATTATCTGGG GCCTTGGCGT CGCCATGGCC ATCTACCTGA CGGCCGGTGT CTCCGGCGCG 11820
 CACCTAAATC CGGCGGTGAC CATTGCCCTG TGGCTGTTCG CCTGTTTGA ACGCCGCAAG 11880
 GTGCTGCCGT TTATTGTTGC CCAGACGGCC GGGGCCTTCT GCGCCGCCGC GCTGGTGTAT 11940
 GGGCTCTATC GCCAGCTGTT TCTCGATCTT GAACAGAGTC AGCATATCGT GCGCGGCACT 12000
 GCCGCCAGTC TTAACCTGGC CGGGGTCTTT TCCACGTACC CGCATCCACA TATCACTTTT 12060
 ATACAAGCGT TTGCCGTGGA GACCACCATC ACGGCAATCC TGATGGCGAT GATCATGGCC 12120
 CTGACCGACG ACGGCAACGG AATTC 12145

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 94 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AGCTTAGGAG TCTAGAATAT TGAGCTCGAA TTCCCGGGCA TGCGGTACCG GATCCAGAAA 60
 AAAGCCCGCA CCTGACAGTG CGGGCTTTTT TTTT 94

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 37 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGAATTCAGA TCTCAGCAAT GAGCGAGAAA ACCATGC 37

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid

-- 88 --

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GCTCTAGATT AGCTTCCTTT ACGCAGC

27

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GGCCAAGCTT AAGGAGGTTA ATTAAATGAA AAG

33

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GCTCTAGATT ATTCAATGGT GTCGGG

26

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GCGCCGTCTA GAATTATGAG CTATCGTATG TTTGATTATC TG

42

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid

-- 89 --

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TCTGATACGG GATCCTCAGA ATGCCTGGCG GAAAAT

36

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 51 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GC GCGGATCC AGGAGTCTAG AATTATGGGA TTGACTACTA AACCTCTATC T

51

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GATACGCCCCG GGTTACCATT TCAACAGATC GTCCTT

36

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

TCGACGAATT CAGGAGGA

18

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid

-- 90 --

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CTAGTCCTCC TGAATTCG

18

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CTAGTAAGGA GGACAATTC

19

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CATGGAATTG TCCTCCTTA

19

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 271 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPP1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Met Lys Arg Phe Asn Val Leu Lys Tyr Ile Arg Thr Thr Lys Ala Asn
1 5 10 15

Ile Gln Thr Ile Ala Met Pro Leu Thr Thr Lys Pro Leu Ser Leu Lys

-- 91 --

20	25	30
Ile Asn Ala Ala Leu Phe Asp Val Asp Gly Thr Ile Ile Ile Ser Gln 35 40 45		
Pro Ala Ile Ala Ala Phe Trp Arg Asp Phe Gly Lys Asp Lys Pro Tyr 50 55 60		
Phe Asp Ala Glu His Val Ile His Ile Ser His Gly Trp Arg Thr Tyr 65 70 75 80		
Asp Ala Ile Ala Lys Phe Ala Pro Asp Phe Ala Asp Glu Glu Tyr Val 85 90 95		
Asn Lys Leu Glu Gly Glu Ile Pro Glu Lys Tyr Gly Glu His Ser Ile 100 105 110		
Glu Val Pro Gly Ala Val Lys Leu Cys Asn Ala Leu Asn Ala Leu Pro 115 120 125		
Lys Glu Lys Trp Ala Val Ala Thr Ser Gly Thr Arg Asp Met Ala Lys 130 135 140		
Lys Trp Phe Asp Ile Leu Lys Ile Lys Arg Pro Glu Tyr Phe Ile Thr 145 150 155 160		
Ala Asn Asp Val Lys Gln Gly Lys Pro His Pro Glu Pro Tyr Leu Lys 165 170 175		
Gly Arg Asn Gly Leu Gly Phe Pro Ile Asn Glu Gln Asp Pro Ser Lys 180 185 190		
Ser Lys Val Val Val Phe Glu Asp Ala Pro Ala Gly Ile Ala Ala Gly 195 200 205		
Lys Ala Ala Gly Cys Lys Ile Val Gly Ile Ala Thr Thr Phe Asp Leu 210 215 220		
Asp Phe Leu Lys Glu Lys Gly Cys Asp Ile Ile Val Lys Asn His Glu 225 230 235 240		
Ser Ile Arg Val Gly Glu Tyr Asn Ala Glu Thr Asp Glu Val Glu Leu 245 250 255		
Ile Phe Asp Asp Tyr Leu Tyr Ala Lys Asp Asp Leu Leu Lys Trp 260 265 270		

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 555 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

-- 92 --

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: DHAB1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```

Met Lys Arg Ser Lys Arg Phe Ala Val Leu Ala Gln Arg Pro Val Asn
1           5           10           15

Gln Asp Gly Leu Ile Gly Glu Trp Pro Glu Glu Gly Leu Ile Ala Met
20           25           30

Asp Ser Pro Phe Asp Pro Val Ser Ser Val Lys Val Asp Asn Gly Leu
35           40           45

Ile Val Glu Leu Asp Gly Lys Arg Arg Asp Gln Phe Asp Met Ile Asp
50           55           60

Arg Phe Ile Ala Asp Tyr Ala Ile Asn Val Glu Arg Thr Glu Gln Ala
65           70           75           80

Met Arg Leu Glu Ala Val Glu Ile Ala Arg Met Leu Val Asp Ile His
85           90           95

Val Ser Arg Glu Glu Ile Ile Ala Ile Thr Thr Ala Ile Thr Pro Ala
100          105          110

Lys Ala Val Glu Val Met Ala Gln Met Asn Val Val Glu Met Met Met
115          120          125

Ala Leu Gln Lys Met Arg Ala Arg Arg Thr Pro Ser Asn Gln Cys His
130          135          140

Val Thr Asn Leu Lys Asp Asn Pro Val Gln Ile Ala Ala Asp Ala Ala
145          150          155          160

Glu Ala Gly Ile Arg Gly Phe Ser Glu Gln Glu Thr Thr Val Gly Ile
165          170          175

Ala Arg Tyr Ala Pro Phe Asn Ala Leu Ala Leu Leu Val Gly Ser Gln
180          185          190

Cys Gly Arg Pro Gly Val Leu Thr Gln Cys Ser Val Glu Glu Ala Thr
195          200          205

Glu Leu Glu Leu Gly Met Arg Gly Leu Thr Ser Tyr Ala Glu Thr Val
210          215          220

Ser Val Tyr Gly Thr Glu Ala Val Phe Thr Asp Gly Asp Asp Thr Pro
225          230          235          240

Trp Ser Lys Ala Phe Leu Ala Ser Ala Tyr Ala Ser Arg Gly Leu Lys
245          250          255

Met Arg Tyr Thr Ser Gly Thr Gly Ser Glu Ala Leu Met Gly Tyr Ser
260          265          270

```

-- 93 --

Glu Ser Lys Ser Met Leu Tyr Leu Glu Ser Arg Cys Ile Phe Ile Thr
 275 280 285
 Lys Gly Ala Gly Val Gln Gly Leu Gln Asn Gly Ala Val Ser Cys Ile
 290 295 300
 Gly Met Thr Gly Ala Val Pro Ser Gly Ile Arg Ala Val Leu Ala Glu
 305 310 315 320
 Asn Leu Ile Ala Ser Met Leu Asp Leu Glu Val Ala Ser Ala Asn Asp
 325 330 335
 Gln Thr Phe Ser His Ser Asp Ile Arg Arg Thr Ala Arg Thr Leu Met
 340 345 350
 Gln Met Leu Pro Gly Thr Asp Phe Ile Phe Ser Gly Tyr Ser Ala Val
 355 360 365
 Pro Asn Tyr Asp Asn Met Phe Ala Gly Ser Asn Phe Asp Ala Glu Asp
 370 375 380
 Phe Asp Asp Tyr Asn Ile Leu Gln Arg Asp Leu Met Val Asp Gly Gly
 385 390 395 400
 Leu Arg Pro Val Thr Glu Ala Glu Thr Ile Ala Ile Arg Gln Lys Ala
 405 410 415
 Ala Arg Ala Ile Gln Ala Val Phe Arg Glu Leu Gly Leu Pro Pro Ile
 420 425 430
 Ala Asp Glu Glu Val Glu Ala Ala Thr Tyr Ala His Gly Ser Asn Glu
 435 440 445
 Met Pro Pro Arg Asn Val Val Glu Asp Leu Ser Ala Val Glu Glu Met
 450 455 460
 Met Lys Arg Asn Ile Thr Gly Leu Asp Ile Val Gly Ala Leu Ser Arg
 465 470 475 480
 Ser Gly Phe Glu Asp Ile Ala Ser Asn Ile Leu Asn Met Leu Arg Gln
 485 490 495
 Arg Val Thr Gly Asp Tyr Leu Gln Thr Ser Ala Ile Leu Asp Arg Gln
 500 505 510
 Phe Glu Val Val Ser Ala Val Asn Asp Ile Asn Asp Tyr Gln Gly Pro
 515 520 525
 Gly Thr Gly Tyr Arg Ile Ser Ala Glu Arg Trp Ala Glu Ile Lys Asn
 530 535 540
 Ile Pro Gly Val Val Gln Pro Asp Thr Ile Glu
 545 550 555

-- 94 --

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 194 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: DHAB2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```

Met Gln Gln Thr Thr Gln Ile Gln Pro Ser Phe Thr Leu Lys Thr Arg
1          5          10          15

Glu Gly Gly Val Ala Ser Ala Asp Glu Arg Ala Asp Glu Val Val Ile
20          25          30

Gly Val Gly Pro Ala Phe Asp Lys His Gln His His Thr Leu Ile Asp
35          40          45

Met Pro His Gly Ala Ile Leu Lys Glu Leu Ile Ala Gly Val Glu Glu
50          55          60

Glu Gly Leu His Ala Arg Val Val Arg Ile Leu Arg Thr Ser Asp Val
65          70          75          80

Ser Phe Met Ala Trp Asp Ala Ala Asn Leu Ser Gly Ser Gly Ile Gly
85          90          95

Ile Gly Ile Gln Ser Lys Gly Thr Thr Val Ile His Gln Arg Asp Leu
100         105         110

Leu Pro Leu Ser Asn Leu Glu Leu Phe Ser Gln Ala Pro Leu Leu Thr
115         120         125

Leu Glu Thr Tyr Arg Gln Ile Gly Lys Asn Ala Ala Arg Tyr Ala Arg
130         135         140

Lys Glu Ser Pro Ser Pro Val Pro Val Val Asn Asp Gln Met Val Arg
145         150         155         160

Pro Lys Phe Met Ala Lys Ala Ala Leu Phe His Ile Lys Glu Thr Lys
165         170         175

His Val Val Gln Asp Ala Glu Pro Val Thr Leu His Ile Asp Leu Val
180         185         190

Arg Glu

```

-- 95 --

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 140 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: DHAB3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Met	Ser	Glu	Lys	Thr	Met	Arg	Val	Gln	Asp	Tyr	Pro	Leu	Ala	Thr	Arg	
1				5					10						15	
Cys	Pro	Glu	His	Ile	Leu	Thr	Pro	Thr	Gly	Lys	Pro	Leu	Thr	Asp	Ile	
			20					25						30		
Thr	Leu	Glu	Lys	Val	Leu	Ser	Gly	Glu	Val	Gly	Pro	Gln	Asp	Val	Arg	
			35				40					45				
Ile	Ser	Arg	Gln	Thr	Leu	Glu	Tyr	Gln	Ala	Gln	Ile	Ala	Glu	Gln	Met	
			50				55				60					
Gln	His	Ala	Val	Ala	Arg	Asn	Phe	Arg	Arg	Ala	Ala	Glu	Leu	Ile	Ala	
65					70					75					80	
Ile	Pro	Asp	Glu	Arg	Ile	Leu	Ala	Ile	Tyr	Asn	Ala	Leu	Arg	Pro	Phe	
				85					90					95		
Arg	Ser	Ser	Gln	Ala	Glu	Leu	Leu	Ala	Ile	Ala	Asp	Glu	Leu	Glu	His	
			100					105						110		
Thr	Trp	His	Ala	Thr	Val	Asn	Ala	Ala	Phe	Val	Arg	Glu	Ser	Ala	Glu	
			115				120						125			
Val	Tyr	Gln	Gln	Arg	His	Lys	Leu	Arg	Lys	Gly	Ser					
			130				135				140					

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 387 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: DHAT

-- 96 --

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Met Ser Tyr Arg Met Phe Asp Tyr Leu Val Pro Asn Val Asn Phe Phe
 1 5 10 15
 Gly Pro Asn Ala Ile Ser Val Val Gly Glu Arg Cys Gln Leu Leu Gly
 20 25 30
 Gly Lys Lys Ala Leu Leu Val Thr Asp Lys Gly Leu Arg Ala Ile Lys
 35 40 45
 Asp Gly Ala Val Asp Lys Thr Leu His Tyr Leu Arg Glu Ala Gly Ile
 50 55 60
 Glu Val Ala Ile Phe Asp Gly Val Glu Pro Asn Pro Lys Asp Thr Asn
 65 70 75 80
 Val Arg Asp Gly Leu Ala Val Phe Arg Arg Glu Gln Cys Asp Ile Ile
 85 90 95
 Val Thr Val Gly Gly Gly Ser Pro His Asp Cys Gly Lys Gly Ile Gly
 100 105 110
 Ile Ala Ala Thr His Glu Gly Asp Leu Tyr Gln Tyr Ala Gly Ile Glu
 115 120 125
 Thr Leu Thr Asn Pro Leu Pro Pro Ile Val Ala Val Asn Thr Thr Ala
 130 135 140
 Gly Thr Ala Ser Glu Val Thr Arg His Cys Val Leu Thr Asn Thr Glu
 145 150 155 160
 Thr Lys Val Lys Phe Val Ile Val Ser Trp Arg Lys Leu Pro Ser Val
 165 170 175
 Ser Ile Asn Asp Pro Leu Leu Met Ile Gly Lys Pro Ala Ala Leu Thr
 180 185 190
 Ala Ala Thr Gly Met Asp Ala Leu Thr His Ala Val Glu Ala Tyr Ile
 195 200 205
 Ser Lys Asp Ala Asn Pro Val Thr Asp Ala Ala Ala Met Gln Ala Ile
 210 215 220
 Arg Leu Ile Ala Arg Asn Leu Arg Gln Ala Val Ala Leu Gly Ser Asn
 225 230 235 240
 Leu Gln Ala Arg Glu Asn Met Ala Tyr Ala Ser Leu Leu Ala Gly Met
 245 250 255
 Ala Phe Asn Asn Ala Asn Leu Gly Tyr Val His Ala Met Ala His Gln
 260 265 270
 Leu Gly Gly Leu Tyr Asp Met Pro His Gly Val Ala Asn Ala Val Leu
 275 280 285

-- 97 --

Leu Pro His Val Ala Arg Tyr Asn Leu Ile Ala Asn Pro Glu Lys Phe
 290 295 300

Ala Asp Ile Ala Glu Leu Met Gly Glu Asn Ile Thr Gly Leu Ser Thr
 305 310 315 320

Leu Asp Ala Ala Glu Lys Ala Ile Ala Ala Ile Thr Arg Leu Ser Met
 325 330 335

Asp Ile Gly Ile Pro Gln His Leu Arg Asp Leu Gly Val Lys Glu Ala
 340 345 350

Asp Phe Pro Tyr Met Ala Glu Met Ala Leu Lys Asp Gly Asn Ala Phe
 355 360 365

Ser Asn Pro Arg Lys Gly Asn Glu Gln Glu Ile Ala Ala Ile Phe Arg
 370 375 380

Gln Ala Phe
 385

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GCGAATTCAT GAGCTATCGT ATGTTTG

27

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GCGAATTCAG AATGCCTGGC GGAAAATC

28

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

-- 98 --

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GGGAATTCAT GAGCGAGAAA ACCATGCG

28

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GCGAATTCTT AGCTTCCTTT ACGCAGC

27

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GCGAATTCAT GCAACAGACA ACCCAAATTC

30

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GCGAATTCAC TCCCTTACTA AGTCG

25

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

-- 99 --

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GGGAATTCAT GAAAAGATCA AAACGATTTG

30

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GCGAATTCTT ATTCAATGGT GTCGGGCTG

29

(2) INFORMATION FOR SEQ ID NO:46

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TTGATAATAT AACCATGGCT GCTGCTGCTG ATAG

34

(2) INFORMATION FOR SEQ ID NO:47

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

GTATGATATG TTATCTTGGA TCCAATAAAT CTAATCTTC

39

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

-- 100 --

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CATGACTAGT AAGGAGGACA ATTC

24

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CATGGAATTG TCCTCCTTAC TAGT

24